Compressive MRI quantification using convex spatiotemporal priors and deep encoder-decoder networks

Medical Image Analysis (2020)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Compressive MRI quantification using convex spatiotemporal priors and deep encoder-decoder networks

Mohammad Golbabaee[a,*], Guido Buonincontri[b], Carolin M. Pirkl[c,d], Marion I. Menzel[c,d], Bjoern H. Menze[c], Mike Davies[e], Pedro A. Gómez[c]

[a] *Computer Science Department, University of Bath, United Kingdom*
[b] *Imago7 Foundation, Pisa, Italy*
[c] *Computer Science Department, Technical University of Munich, Germany*
[d] *GE Healthcare, Munich, Germany*
[e] *School of Engineering, University of Edinburgh, United Kingdom*

ARTICLE INFO

Article history:

Keywords: Magnetic Resonance Fingerprinting, compressed sensing, convex model-based reconstruction, residual network, encoder-decoder network

ABSTRACT

We propose a dictionary-matching-free pipeline for multi-parametric quantitative MRI image computing. Our approach has two stages based on compressed sensing reconstruction and deep learned quantitative inference. The reconstruction phase is convex and incorporates efficient spatiotemporal regularisations within an accelerated iterative shrinkage algorithm. This minimises the under-sampling (aliasing) artefacts from aggressively short scan times. The learned quantitative inference phase is purely trained on physical simulations (Bloch equations) that are flexible for producing rich training samples. We propose a deep and compact encoder-decoder network with residual blocks in order to embed Bloch manifold projections through multi-scale piecewise affine approximations, and to replace the non-scalable dictionary-matching baseline. Tested on a number of datasets we demonstrate effectiveness of the proposed scheme for recovering accurate and consistent quantitative information from novel and aggressively subsampled 2D/3D quantitative MRI acquisition protocols.

1. Introduction

Quantification of the intrinsic nuclear magnetic resonance (NMR) characteristics (Tofts, 2005) has proven powerful for tissue identification and tracking pathological changes. Despite many potentials, standard quantitative MRI (QMRI) approaches have very long acquisition times and for this reason, are not widely applicable in clinical setups. Magnetic Resonance Fingerprinting (MRF) has emerged to overcome this challenge (Ma et al., 2013). MRF uses short excitation sequences capable of simultaneously encoding multitudes of NMR properties and further adopts Compressed Sensing (CS) to subsample a tiny fraction of the spatiotemporal k-space information (Jiang Y et al., 2015; Rieger et al., 2017; Wright et al., 2018; Jiang et al., 2017; Lustig et al., 2007). Estimating the underlying quantitative maps therefore becomes a highly ill-posed inverse problem.

Baseline approaches to solve the MRF inverse problem rely on *dictionary matching* (DM), primarily for parameter inference i.e. estimating quantitative maps from back-projected images, or further for promoting temporal-domain priors within model-based MRF reconstructions (Davies et al., 2014). However DM's complexity (storage/runtime) does not scale well to the emerging multi-parametric QMRI applications. *Deep learning* MRF approaches recently emerged to address this issue (Hoppe et al., 2017; Cohen et al., 2018; Virtue et al., 2017; Golbabaee et al., 2019a). Back-projected images are fed into

*Corresponding author e-mail: m.golbabaee@bath.ac.uk

a compact neural network that temporally processes voxel-wise MRF signal evolutions, so-called fingerprints, and replaces DM for quantitative inference. Trained with independently corrupted noisy fingerprints, such networks are unable to correct for dominant spatially-correlated (aliasing) artefacts appearing in heavily undersampled aquisitions. While larger convolutional models (Balsiger et al., 2018; Fang et al., 2019a) capture spatiotemporal information to resolve aliasing artefacts, labelled QMRI datasets (i.e. ground truth multi-parametric anatomical maps) that are necessary to train these models particularly in novel applications are scarce and hence place adaption of these models at the risk of overfitted predictions. Further, current approaches along this line build customised de-noisers (de-aliasing) and require expensive re-training by changing sampling parameters i.e. the forward model.

This work aims to address these shortcomings through a two-stage DM-free pipeline: First, we take a CS approach to *spatiotemporally process* the k-space data and minimise undersampling artefacts in the reconstructed image time-series, and second we feed the resulted sequence to a deep and compact *encoder-decoder* network with residual blocks for per-voxel quantitative inference. We cast our spatiotemporally regularised reconstruction as a *convex* optimisation problem which enjoys reproducible global solutions regardless of initialisation and can be implemented with a momentum-accelerated algorithm with fast convergence. We further provide geometrical insights to the mechanism behind the proposed deep inference approach. We show that the network provides a *multi-resolution piecewise affine approximation* to the Bloch response manifold projection. Rather than memorising a large MRF dictionary, the network hierarchically clusters this manifold through deep layers and learns a compact set of *deep regressing filters* for parameter inference. The proposed pipeline is validated on a number of experiments using a novel multi-parametric acquisition sequence for *2D and 3D quantitative brain imaging*. Our approach can flexibly apply and report consistent predictions for different k-space readouts and further outperforms shallow learned inference models related to the Gaussian kernel fitting.

Paper organisation: We review related works in section 2. Section 3 presents the inverse imaging problem model. Section 4 presents our reconstruction and quantitive inference pipeline. Section 5 presents our geometrical insight to the network's performance for deep quantitative inference. In Section 6 we present and discuss our experimental results, and finally we conclude in section 7.

Notations: Throughout $\|.\|$ denotes the Euclidean norm of a vector or a matrix, $\|.\|_{TV}$ denotes the Total Variation (TV) of a 2D or 3D spatial image defined by the sums of its gradient magnitudes (Rudin et al., 1992). Matrix rows and columns are denoted by $X_{(i,.)}$ and X_i respectively.

2. Related works

Multi-parametric quantification based on fingerprinting, DM and its low-rank singular value decomposition (SVD) compressed variant were proposed in (Ma et al., 2013; McGivney et al., 2014). Reconstructing image time-series from k-space data was non-iterative and used zero-filling (ZF). Inspired by CS, later studies adopted model-based reconstructions to reduce subsampling (aliasing) artefacts and to pave the path for aggressively shorter scan times (Davies et al., 2014; Assländer et al., 2018). These methods are based on non-convex optimisation (iterative) without momentum-acceleration, and require DM per iteration in order to promote temporal-domain Bloch dynamic priors. While fast search schemes (Cauley et al., 2015; Cline et al., 2017; Golbabaee et al., 2019b) could partly improve the runtime, DM remains a computational/storage bottleneck for multi-dimensional imaging problems involving multi-parametric dictionaries. Further, for some k-space subsampling patterns, including those adopted in our experiments, using only a temporal-domain prior is insufficient to produce artefact-free reconstructions (see e.g. (Cline et al., 2017; Golbabaee et al., 2019b)). Spatial domain regularisations were integrated into DM (Cline et al., 2017; Arberet et al., 2019; Gómez et al., 2015, 2016), however these methods require costly DM per iteration, are nonconvex and without momentum-acceleration. DM-free convex reconstructions based on low-rank priors were proposed, some (Zhao et al., 2018; Mazor et al., 2018; Song et al., 2019) with no spatial regularisations hence prone to artefacts in highly ill-posed problems, some (Mazor et al., 2018; Song et al., 2019) without temporal dimensionality reduction and long runtimes, and some using patch-based spatiotemporal low-rank regularisation (Bustin et al., 2019; Jaubert et al., 2020) but encountering long runtimes due to non-accelerated iterations and per-iteration costly SVD decompositions. Our work mitigates these issues: we propose an alternative convex formulation for the MRF reconstruction problem (our preliminary results appeared in (Golbabaee et al., 2019c)). We add spatial TV regularisation for the dimension-reduced image time-series while enforcing temporal-domain priors through a (low-rank) subspace representation of the dictionary instead of DM. This optimisation can be solved with momentum-accelerated iterative algorithms with fast global convergence.

On the other hand, deep learning MRF approaches recently emerged to address the non-scalability of DM. Many works use non-iterative baselines (Ma et al., 2013; McGivney et al., 2014) for reconstruction, and for quantitative inference they replace DM with a neural network. These methods broadly divide in two camps: the first group learns temporal-domain dynamics from simulating Bloch equations; hence is rich with training data (see e.g. (Cohen et al., 2018; Virtue et al., 2017; Golbabaee et al., 2019a; Oksuz et al., 2019) and also a kernel machine approach for shallow learning (Nataraj et al., 2018)). The second group use convolutional layers to also learn spatial domain regularities, see e.g. (Hoppe et al., 2017, 2019; Balsiger et al., 2018, 2019; Fang et al., 2019a,b), but they require training on ground truth quantitative anatomical maps that may not be largely available as for the mainstream qualitative MRI. Our quantitative inference approach belongs to the first camp. We provide a geometrical interpretation for our deep inference approach. Importantly, we replace ZF by our DM-free spatiotemporally regularised (model-based) reconstruction to remove undersampling artefacts before being fed to the network. This enables aggressively short-time 2D/3D quantitative imaging pro-

tocols produce consistent results using a fast and non memory-intensive computational pipeline.

3. Compressive QMRI acquisition model

The compressed sensing approach adopted by MRF for acquiring quantitative information follows a linear spatiotemporal model (Ma et al., 2013):

$$Y = \mathcal{A}(\overline{X}) + \xi, \tag{1}$$

where $Y \in \mathbb{C}^{T\times m}$ is the multi-coil k-space measurements collected at $t = 1, \dots, T$ temporal frames and corrupted by some noise ξ. The Time-Series of Magnetisation Images (TSMI) —to be reconstructed—is an image sequence represented by a complex-valued matrix $\overline{X}$ of spatiotemporal resolution $T \times n$ i.e. n spatial voxels across T temporal frames. The forward operator $\mathcal{A} := F_\Omega S$ models the multi-coil sensitivities operator S, and the Fourier transform F subsampled according to a set of *temporally-varying* k-space locations Ω.

The tissues' quantitative properties in each voxel are encoded in a temporal signal at the corresponding column of the TSMI matrix. This signal records the magnetisation response of proton dipoles to dynamic *excitations* in the form a sequence of flip angles FA (magnetic field rotations) applying with certain repetition (TR) and echo (TE) times. Tissues with different NMR characteristics respond distinctively to excitations. QMRI/MRF rely on this principle to estimate quantitative characteristics from the (computed) TSMI. Per-voxel v magnetisation responses of the TSMI scaled by the *proton density* γ_v are modelled as

$$\overline{X}_v \approx \gamma_v \mathcal{B}(\Theta_v; \text{FA, TR, TE}) \quad \forall v \in 1, \dots, n \tag{2}$$

where the *Bloch response* (in short-hand notation) $\mathcal{B}(\Theta_v) : \mathbb{R}^p \to \mathbb{C}^T$ is a nonlinear mapping from per-voxel intrinsic NMR properties Θ_v to the corresponding (discrete-time) solution of the *Bloch differential equations* which captures the overall transient-state macroscopic dynamics of a voxel (Jaynes, 1955). Our experiments use sequences that simultaneously encode $p = 2$ characteristics in each voxel i.e. the longitudinal T1 and transverse T2 relaxation times. This could be further extended to include other properties e.g. off resonance frequencies, $T2^*$, diffusion and perfusion (Rieger et al., 2017; Wright et al., 2018; Jiang et al., 2017).

3.1. Low-dimensional manifold and subspace models

Estimating Θ (i.e. quantification) requires long enough sequences $T > p$ to create contrast between different tissues' responses. As such the Bloch responses despite their high ambient dimension live on a *low p-dimensional (nonlinear) submanifold* of $\mathbb{C}^T$. Further it is observed that for certain excitation sequences, including those used in our experiments, this manifold is approximately embedded in a *low-rank subspace* $\text{Range}(V) \subset \mathbb{C}^T$ represented by an orthonormal matrix $V \in \mathbb{C}^{T\times s}$ where $p < s \ll T$. Hence the following dimension-reduced alternatives for models (1) and (2) can be deduced:

$$Y \approx \mathcal{A}(VX) \tag{3}$$

$$X_v \approx \gamma_v V^H \mathcal{B}(\Theta_v) \tag{4}$$

where $X \in \mathbb{C}^{s\times n}$ is the dimension-reduced TSMI. This compact representation is the basis for the subspace compression methods (Assländer et al., 2018; Zhao et al., 2018) and is proven beneficial to the runtime and accuracy (by noise trimming) of the reconstructions.

3.2. Model-fitting for parameter inference

Fitting computed TSMIs to the Bloch response model is central to QMRI. Per-voxel model-fitting according to (4) for obtaining the NMR characteristics and proton density reads (see e.g. (McGivney et al., 2014; Davies et al., 2014)):

$$\widehat{\Theta}_v = \mathcal{P}_{\mathcal{B}}(\widehat{X}_v) := \text{argmin}_\Theta \|\widehat{X}_v - V^H \mathcal{B}(\Theta)\| \tag{5}$$

$$\widehat{\gamma}_v = \langle \widehat{X}_v, V^H \mathcal{B}(\widehat{\Theta}_v)\rangle \tag{6}$$

We assumed without losing generality having normalised Bloch responses. We refer to $\mathcal{P}_{\mathcal{B}}(.)$ as the *Bloch response manifold projection.* This projection is nonconvex and oftentimes intractable for the generally complicated Bloch responses adopted by the MRF sequences. The MRF framework instead approximates (5) by *dictionary matching* (DM). A *fingerprint* dictionary $D = \{D_j\}$ is constructed for sampling the manifold of Bloch responses through a fine-grid discretesation of the parameter space $[\Theta] = [T1] \times [T2] \times \dots$ and exhaustively simulating the Bloch responses $D_j := \mathcal{B}([\Theta_j])$ for all combinations of the quantised parameters. The DM step identifies the most correlated fingerprint (and the underlying NMR parameters) for each voxel of the reconstructed TSMI:

$$\mathcal{P}_{\mathcal{B}}(\widehat{X}_v) \approx \text{argmin}_j \|\widehat{X}_v - V^H D_j\| \tag{7}$$

through a nearest neighbour *search* that is itself a projection onto the discrete set of fingerprints i.e. a *point-wise approximation* to the (continuous) Bloch response manifold.

Viewing fingerprints as training samples, the dictionary can be factorised through principal component analysis (PCA) (McGivney et al., 2014):

$$DD^H \approx V\Lambda V^H \tag{8}$$

for unsupervisedly learning the low-rank subspace representation of the Bloch responses. This representation helps to reduce temporal dimension and can be coupled with fast search schemes (Cauley et al., 2015; Golbabaee et al., 2019b; Cline et al., 2017) to accelerate DM runtime. However any form of DM (fast or exhaustive search) remains non-scalable and creates storage overhead in multi-parametric QMRI applications because the number of dictionary atoms *exponentially* grows with p.

4. DM-free image reconstruction and parameter inference pipeline

Our DM-free image computing pipeline consists of two stages: i) reconstructing TSMIs from undersampled k-space measurements and then ii) approximate model-fitting according to (4) for parameter inference. A set of simulated fingerprints (could be MRF dictionary) sample the Bloch response

model and are used only for training (pre-processing) in order to learn three temporal-domain models: i) a dimension-reduced (low-rank) subspace representation for the Bloch responses, ii) an encoder network to map noisy fingerprints to the NMR parameters, and iii) a decoder network to generate clean Bloch responses from the NMR parameters.

4.1. Convex TSMI reconstruction with LRTV algorithm

The (McGivney et al., 2014)'s baseline backprojects k-space data to form dimension-reduced TSMI using the adjoint of $\mathcal{A}V$:

$$\widehat{X} = V^H \mathcal{A}^H(Y) \in \mathbb{C}^{s\times n} \tag{9}$$

prior to the DM inference. Modern QMRI/MRF acquisitions aggressively curtail the scan times by using short excitation sequences and severe spatial (k-space) subsampling. As such the inverse problem (1) becomes highly ill-posed and ZF/backprojection (which is not an inversion) results in aliasing artefacts in the reconstructed TSMI. Errors made at this stage can be indeed significant (see experiment results), they propagate to the parameter inference step and deteriorate the overall quantification accuracy.

To address this issue, we propose the following convex and DM-free optimisation dubbed as Low-Rank and Total-Variation (LRTV) reconstruction to compute dimension-reduced TSMIs using joint spatiotemporal regularisations:

$$\widehat{X} = \text{argmin}_{X\in\mathbb{C}^{s\times n}} \|Y - \mathcal{A}(VX)\|^2 + \sum_{i=1}^{s} \lambda_i \|X_{(i,.)}\|_{TV} \tag{10}$$

The first term minimises discrepancies between the k-space measurements and the solutions through the factorised forward model $Y \approx \mathcal{A}(VX)$. As such LRTV adopts a temporal-domain prior through the subspace model (i.e. the low-rank factorisation $\overline{X} \approx VX$) which provides a compact and convex (in fact linear) *relaxed representation* for the Bloch response model instead of using the MRF dictionary. LRTV additionally adopts Total-Variation (TV) regularisation. Each component of the TSMI corresponds to a spatial 2D or 3D volumetric image $X_{(i,.)}$ (matrix row), where penalising its TV norm promotes spatial-domain regularities via sparse image gradients. $\lambda_i > 0$ control per (subspace) component regularisation levels.

Our LRTV formulation (10) can be efficiently solved using Fast Iterative Shrinkage Algorithm with Nesterov momentum acceleration and backtracking step-size (Beck and Teboulle, 2009; Nesterov, 1983). The pseudo-code is available in supplementary section SI. In nutshell, each iteration k computes:

$$\begin{cases} \nabla = X^k - \mu_k V^H \mathcal{A}^H\Big(\mathcal{A}(VX^k) - Y\Big) \\ Z^k_{(i,.)} = \mathbf{Prox}_{\lambda_i\mu_k}(\nabla_{(i,.)}) \quad \forall i = 1,\dots s \\ X^{k+1} = Z^k + \Big(\frac{k-1}{k+2}\Big)(Z^k - Z^{k-1}) \end{cases} \tag{11}$$

The first and third lines are the gradient and momentum-acceleration updates, respectively. The second line computes a small number $s \ll T$ of shrinkage operations for the 2D/3D images in each subspace component $\mathbf{Prox}_\alpha(x) := \text{argmin}_u \frac{1}{2}\|x-u\|^2 + \alpha\|u\|_{TV}$, which can be efficiently done on a GPU using the Primal-Dual algorithm (Chambolle and Pock, 2011).

With an all-zero initialisation, the first line of (11) recovers ZF (McGivney et al., 2014) in the first iteration. Setting $\lambda = 0$ recovers the LR formulation in (Zhao et al., 2018). LR is a convex relaxed alternative to DM-based models (Davies et al., 2014; Assländer et al., 2018), wherein temporal-only priors based on the MRF dictionary are replaced by the low-rank subspace. Note that the size of V is independent of the number of fingerprints (used for training). Hence the solver does not face a memory bottleneck and the slow progress of computing DM per iteration. While for certain (Cartesian) sampling schemes this temporal model can decently regularise the inversion (Benjamin et al., 2019), for other important sampling patterns e.g. non-cartesian spiral and radial readouts used in our experiments, it turns out to be inadequate and fails to output artefact-free TSMIs (see section 6). Multi-prior CS solvers are proven effective for highly undersampled systems by further restricting degrees of freedom of data (Golbabaee and Vandergheynst, 2012a,b). The LRTV uses this fact by setting $\lambda > 0$ and adding spatial priors to sufficiently regularise the problem. Besides being DM-free, the LRTV has other advantages over its non-convex spatiotemporal alternatives (Cline et al., 2017; Arberet et al., 2019), including a tractable way to incorporate multiple priors[1], momentum-acceleration for fast convergence and reproducible global solutions regardless of initialisation.

4.2. MRFResnet for parameter inference

Instead of using a large-size dictionary for DM, we propose training and using a compact network coined as *MRFResnet* in the form of an encoder-decoder with deep residual blocks, shown in Figure 1. Encoder-decoders family have proven powerful in denoising tasks through creating an *information bottleneck* which corresponds to learning a low-dimensional manifold model for capturing (nonlinear) intrinsic signal structures (Vincent et al., 2010). In our task computed TSMI voxels are processed by such a model to create clean magnetisation responses as well as estimating the intrinsic NMR parameters in a computably efficient manner. The $p = 2$ neurons bottleneck (in Figure 1) has a physical interpretation: fitting noisy temporal trajectories to the nonlinear Bloch model with limited $p \ll T$ degrees of freedom determined by the T1 and T2 quantities.

4.2.1. Encoder

This network learns to approximate Bloch manifold projections through a continuous mapping $\mathcal{R} : \widehat{X}_v \to \Theta_v$ parametrised by the network's weights and biases $\{W,\beta\}$:

$$\mathcal{R}(x) \equiv h^{(N+1)}(x) = \varphi\Big(W^{(N+1)}h^{(N)}(x) + \beta^{(N+1)}\Big) \tag{12}$$

where $h^{(i)}$ the outputs of $i = 1,\dots,N$ residual blocks are

$$h^{(i)}(x) = \varphi\Big(h^{(i-1)}(x) + g^{(i)}(x)\Big),$$
$$\text{and} \quad g^{(i)}(x) = W^{(i,2)}\varphi\Big(W^{(i,1)}h^{(i-1)}(x) + \beta^{(i,1)}\Big) + \beta^{(i,2)},$$

[1] In non-convex (e.g. DM-based) approaches incorporating extra priors such as spatial regularity constraints is not always algorithmically tractable e.g. sequential projections on two sets where one is non-convex may not result in projecting onto the intersection.

$h^{(0)}(x) = x$ and the ReLU activations $\varphi(x) = \max(x, 0)$ are used throughout. The inputs are the normalised temporal voxels of the dimension-reduced TSMI. The network is trained on simulated noisy Bloch responses (see section 6.2) so that the approximate projection holds

$$\mathcal{R}(x) \approx \mathcal{P}_{\mathcal{B}}(x) \tag{13}$$

in a neighbourhood of the (compressed) Bloch manifold.

4.2.2. Decoder

The proton density (PD) is a scaling factor that amplifies the Bloch responses in each voxel. Hence after estimating other nonlinear NMR parameters (e.g. T1/T2) using the encoder part, PD can be explicitly estimated by $\langle \widehat{X}_v, V^H \mathcal{B}(\widehat{\Theta}_v) \rangle$ in (6). This would however require either storing a dense dictionary or evaluating Bloch responses for all voxels' parameters, which can be computationally intensive. Instead we train a *decoder* network $\mathcal{G}(.)$ which for given NMR parameters it approximately generates

$$\mathcal{G}(\Theta_v) \approx V^H \mathcal{B}(\Theta_v) \tag{14}$$

the corresponding compressed Bloch responses (clean fingerprints) in short runtimes. This allows (6) to be easily applied without significant computations. For the sequence design used in our experiments, it turns out that a fully-connected shallow network with one hidden layer and ReLU activations can approximate well this step.[2] Unit dimensions are customised to a sequence used in our experiments encoding T1/T2 relaxation times, with reduced subspace dimension $s = 10$. Encoder has $N = 6$ residual blocks of 10 neurons width, and decoder has 300 neurons in its single hidden layer.

The subspace compression helps reduce model sizes in both networks (hence reducing risk of overfitted predictions) and also reduce required training resources compared to uncompressed deep MRF approaches (Cohen et al., 2018; Virtue et al., 2017). Further to avoid losing discrimination between fingerprints —e.g. by a magnitude-only data processing (Cohen et al., 2018) —we adopt a practical phase-correction step (Golbabaee et al., 2019b; Cline et al., 2017) to align phases of the complex-valued TSMIs and training samples before being fed to the MRFResnet. Complex-valued phases estimated from the first principal components of the dimension-reduced TSMI pixels and the training samples are used for phase-alignment. This treatement allows the network without losing generality have real-valued parameters and approximate real-valued mappings.

5. Hierarchical partitioning of the Bloch response manifold

In this part we show that the MRFResnet provides a multi-scale piecewise affine approximation to the Bloch response manifold projection (5). Hierarchical partitioning and multi-scale approximations are also central to the fast search schemes proposed for the DM-based MRF (see illustrations in (Golbabaee et al., 2019b, 2017)). However unlike any form of DM (fast or exhaustive) that creates point-wise approximations for (7), MRFResnet does not memorise a dictionary and rather uses it to learn and efficiently encode a compact set of partitions and *deep matched-filters* for affine regression of the NMR quantities.

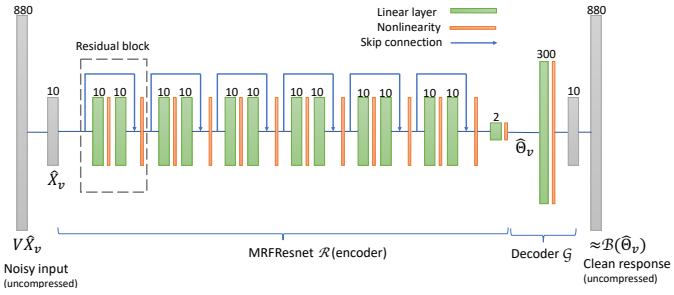


Fig. 1: MRFResnet (encoder) for T1/T2 inference, the Bloch response generative network $\mathcal{G}$ (decoder), and the *implicit* linear dimensionality reduction/expansion (first/last) layers using the subspace model V^H/V.

5.1. Affine spline function approximation

The MRFResnet encoder (also its decoder network) is composed of linear connections and piecewise linear ReLU activations. This results in piecewise affine functions $h^{(i)}(x)$ after each residual block as well as the end-to-end mapping $\mathcal{R}(x)$ (see e.g. (Balestriero and Baraniuk, 2018; Montufar et al., 2014)). Further, $\mathcal{R}$ is *Lipschitz continuous* for continuous activation functions as above and for bounded $\{W^{(i)}, \beta^{(i)}, i\}$.

Theorem 1. *Denote by* $z : \mathbb{R}^s \to \mathbb{R}^p$

$$z(x) := W^{(N+1)} h^{(N)}(x) + \beta^{(N+1)} \tag{15}$$

the weighted outputs *in* (12) *before the last non-linearity.*[3] *The following affine spline representation holds for MRFResnet:*

$$z(x) = \mathbf{A}[x]x + \mathbf{b}[x] := \sum_r (A_r x + b_r)\, \iota_{\Omega_r}(x), \tag{16}$$

where $\iota_{\Omega_r}(x)$ *is the indicator function with respect to a segment (set)* $\Omega_r \in \mathbb{R}^s$*, returning* x *if it belongs to the segment and 0 otherwise —segments form a disjoint partitioning of the input space with affine boundaries. Matrices* $A_r \in \mathbb{R}^{p \times s}$ *and vectors* $b_r \in \mathbb{R}^p$ *define the corresponding slopes and offsets for the input-output affine mapping in each segment. Shorthands* $\mathbf{A}[x] : \mathbb{R}^s \to \mathbb{R}^p$ *and* $\mathbf{b}[x] : \mathbb{R}^s \to \mathbb{R}^p$ *represent the input-dependent (piece-wise affine) mapping of* $z(x)$*.* $\mathbf{b}[x]$ *represents* p *input-dependent offsets. Similarly,* $\mathbf{A}[x]$ *is an input-dependent* $p \times s$ *matrix where each row is a deep matched-filter returning its correlation with* x *for each output.*

Proof can be found in (Balestriero and Baraniuk, 2018) for general feedforward networks with fully-connected, convolutional, pooling and/or residual layers and using any piecewise-linear activations. During training, MRFResnet encoder learns

[2] We also observed a similar network complexity for generating responses to the well-known FISP sequence (Jiang Y et al., 2015). However, we did not achieved accurate predictions using shallow architectures of comparable sizes for $\mathcal{R}$ (two layer were needed at least however with larger model than MRFResnet (Golbabaee et al., 2019a,c)). This suggests that generating clean responses (decoding) was easier than projecting noisy fingerprints to their generative parameters (encoding), and the latter requires deep processing (see section 5).

[3] The last ReLU layer in $\mathcal{R}$ is for imposing the positivity of T1/T2 values, and therefore the prediction task is mainly done by the preceding layers.

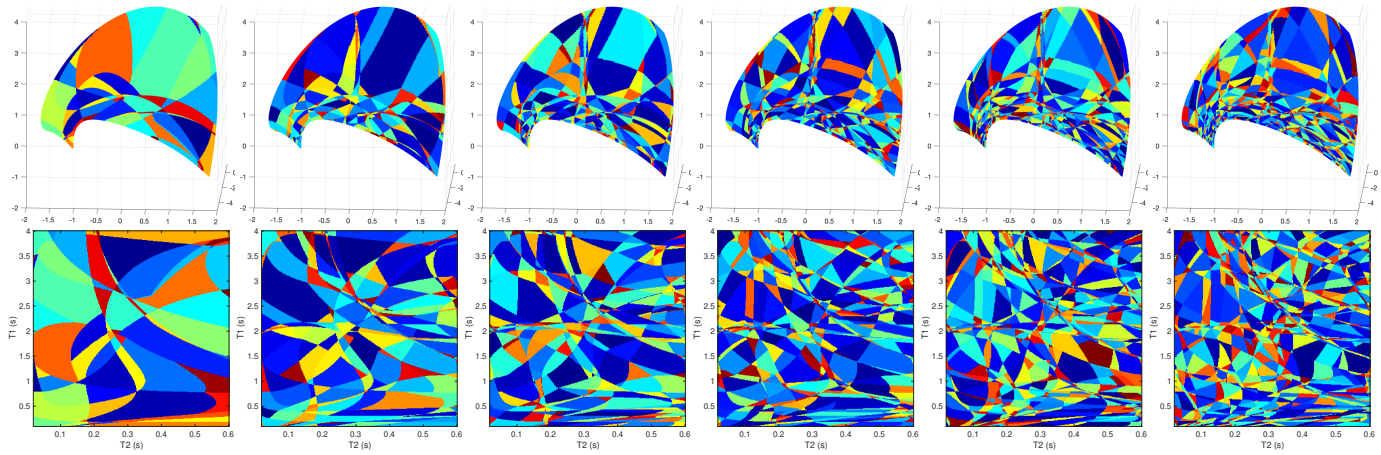
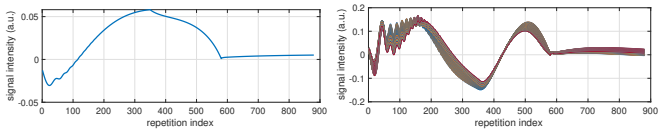
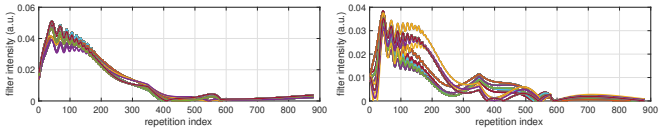


Fig. 2: Coarse-to-fine partitioning of the Bloch manifold (top row—manifold data is visualised using PCA) sampled by a dense fingerprinting dictionary, and their generative T1/T2 parameters (bottom row) using MRFResnet. From left to right figures illustrate learned colour-coded partitions after each residual block.



(a) The average magnetic response

(b) Centred magnetic responses



(c) T1 deep matched filter

(d) T2 deep matched filter

Fig. 3: (a) The mean and (b) centred Bloch responses within the range (T1, T2) $\in [1000-1200]\times[80-110]$ (ms). (c)-(d) The end-to-end match filters learned by the MRFResnet to regress T1/T2 values are shown across the original (non-compressed) temporal dimension.

$\{W^{(i)},\beta^{(i)}\}$ or equivalently $\{\mathbf{A}[x],\mathbf{b}[x]\}$ to provide a continuous and piece-wise affine *approximation* for (5). The universal approximation theorem (Cybenko, 1989) states that a shallow network with one but very wide hidden layer can do this. Deeper networks are however more practical to efficiently reduce the number of hidden units (Delalleau and Bengio, 2011). Indeed, we experimentally observe this (section 6.3) by comparing MRFResnet to a shallow learning scheme related to (Nataraj et al., 2018) based on Kernel Machines (KM) and random features (Wilson et al., 2016).

5.2. *Visualising MRFResnet segments on Bloch manifold*

Remark 1. *Continuity of $z(x)$ implies that adjacent segments $\Omega_r,\Omega_{r'}$ correspond to distinct $A_r, A_{r'}$. Indeed, if $A_r = A_{r'}$ and the only difference is in the offsets $b_r \neq b_{r'}$, then $\Omega_r,\Omega_{r'}$ won't intersect on boundaries. Therefore they are not adjacent segments unless contradicting the continuity assumption.*

This remark gives an idea for visualizing the input space segments. For densely sampled input signals x, we compute derivatives of the weighted outputs (15) with respect to inputs using *back propagation*. These will determine the input-dependant slopes in the affine spline formulation (16) i.e. rows of $\mathbf{A}[x]$ at a point x are populated as follows $\forall\, j = 1,2,\ldots,p$:

$$\mathbf{A}[x]_{(j,.)} = \left[\frac{\partial z_j(x)}{\partial x_1}, \frac{\partial z_j(x)}{\partial x_2}, \ldots, \frac{\partial z_j(x)}{\partial x_s}\right]. \tag{17}$$

By vector quantisation (e.g. k-means clustering) we cluster regions of x that output distinct slopes A_r and identify the segments Ω_r. Similar routine could apply to compute input space partitions by clustering back-propagated output derivates after each residual block (Theorem 1 and Remark 1 also hold for the intermediate blocks of $\mathcal{R}$).

According to (Balestriero and Baraniuk, 2018) as we progress into deeper layers, partitions will be subdivided into smaller segments in a *hierarchal* fashion. This can be observed in Figure 2 where we adopted the above routine for the T1/T2 encoding MRF sequence used in our experiments and visualised multi-scale (from coarse-to-fine) partitions obtained after each residual layer. The Bloch response manifold is sampled across fine-gridded T1/T2 values (i.e. MRF dictionary) to visualise the intersection of the input space segments with this manifold (results are visualised across the three dominant principal component axes). MRFResnet encoder learns about a thousand partitions for its end-to-end mapping $z(x)$. In the light of (16) we know that for each partition Ω_r the network implicitly encodes $p = 2$ deep matched-filters (the rows of $\mathbf{A}[x]$ or alternatively A_r) and an offset term to locally linearly regress the T1/T2 outputs in that segment. As such instead of memorising >100K dictionary atoms used for training, the network learns a compact piece-wise affine approximation to the Bloch manifold projection (5) as a rapid and memory-efficient alternative to DM's point-wise approximation (7). The total number of parameters used by the MRFResnet (Table 1) are two hundreds times less than the size of the dimension-reduced MRF dictionary. Figure 3 shows the Bloch responses for a range of T1/T2 values, as well as deep matched-filters learned by MRFResnet to predict each of these quantities in this range from noisy inputs. Computed through (17), match-filters are one-dimensional analogues of the *saliency maps* a.k.a. *deep dream images* (Simonyan and Zisserman, 2014), measuring sensitivities of the T1/T2 output neurons with respect to the inputs.

6. Numerical experiments and discussions

In the spirit of reproducible research, the source codes of the proposed algorithms are available at `https://github.com/mgolbabaee/LRTV-MRFResnet-for-MRFingerprinting`.

6.1. Datasets and 2D/3D acquisition parameters

Methods are tested on the Brainweb *in-silico* phantom (see supplementary materials), a EUROSPIN TO5 phantom (*in-vitro*) (Lerski and De Certaines, 1993), and *in-vivo* acquisitions of a healthy volunteer and a glioma patient. *In-vitro* and *in-vivo* volunteer data were acquired on a 1.5T GE HDxT scanner using 8-channel receive-only head RF coil. For glioma patient, data were acquired on a 3T GE MR750 system. In both cases, the excitation sequence has $T = 880$ repetitions and jointly encodes T1/T2 values using an inversion pulse followed by a flip angle schedule that linearly ramps up from 1° to 70° in repetitions 1-400, ramps down to 1° in repetitions 400-600, and then stays constant to 1° for repetitions 600-880 (see details in (Gómez et al., 2020)). Three non-Cartesian readout trajectories were tested: 2D/3D variable density spiral and 2D radial k-space subsampling patterns. Throughout we used inversion time=18 ms, fixed TR=12/7.8ms for the volunteer/patient acquisitions, and TE = 0.46/2.08 ms for spiral/radial acquisitions, respectively. For the 2D/3D acquisitions we had $200^2/200^3$ (mm^2/mm^3) FOV and $200^2/200^3$ voxels image/tensor size, respectively. Further, the total number of interleaves for the 2D/3D spiral and 2D radial readouts were 377/48'400 and 967, respectively. Only one radial spoke (or spiral arm) was sampled at each of the 880 timeframes, resulting in aggressive acceleration (undersampling) factors 62×, 252× and 262× with respect to a fully sampled 2D spiral, 2D radial and 3D spiral acquisitions, correspondingly. The total acquisition times for the 2D scans were 10:56 seconds, and for the 3D volunteer/patient scans were 9:51/6:25 minutes, respectively. For all cases coil sensitivities were computed from undersampled data using an adaptive coil combination scheme (Walsh et al., 2000).

6.2. Tested algorithms

For TSMI reconstruction we compare model-based iterative methods: LRTV, LR (Zhao et al., 2018) and FLOR (Mazor et al., 2018) which are convex and DM-free, and the non-convex DM-based AIR-MRF (Cline et al., 2017). LRTV and LR solve (10) with spatiotemporal ($\lambda > 0$) and temporal-only ($\lambda = 0$) regularisations, respectively. Note that the latter solves the same problem as (Zhao et al., 2018) but with Nesterov acceleration. Here, AIR-MRF uses Gaussian low-pass filtering (to avoid Gibbs artefacts) to regularise the spatial domain. It further uses MATLAB's in-build kd-tree searches for fast DM per-iteration. The maximum number of iterations for these algorithms were set to 30, but each stops earlier if the relative change in their objectives is below tolerance 10^{-4} (convergence). Further we compare against reconstruction non-iterative baselines ZF (9) (McGivney et al., 2014) and View-Sharing (VS) (Buonincontri and Sawiak, 2016). VS aggregates spatial k-space data within neighbouring temporal frames to increase per-frame samples and enhance spatial resolutions in a non model-based fashion. For quantitative inference and besides DM baseline, we compare various deep learning MRF models including the proposed MRFResnet, and also a shallow learning method based on Gaussian Kernel Machines (KM) related to (Nataraj et al., 2018).[4]

	Total # params.	T1 (ms) MAE	T1 (%) MAPE	T2 (ms) MAE	T2 (%) MAPE	$\mathcal{B}$ (%) NRMSE
MRFResnet	5.2k	7.19	0.91	1.91	1.05	0.86
KM fitting	44k	29.93	3.97	15.84	9.68	12.88

Table 1: Prediction performances of MRFResnet and KM.

6.2.1. Learned models

Except FLOR which does not use dimensionality reduction, other methods use a $s = 10$ dimensional subspace model a-priori learned from Bloch response simulations following the SVD analysis of (McGivney et al., 2014) (the NRMSE difference between the subspace and original dictionaries were less than 7.16×10^{-7}). For this, a dictionary of $d = 113781$ atoms sampling the T1=[100:10:4000] (ms) and T2=[20:2:600] (ms) grid was simulated using the Extended Phase Graph formalism (Weigel, 2015). This subspace-compressed dictionary was directly used in DM, whereas for learning-based inference it was only used for model training and validation. Clean fingerprints were used for training MRFResnet decoder $\mathcal{G}$ i.e. Bloch response generation network. Noisy fingerprints each corrupted with i.i.d. noise $\sim \mathcal{N}(0, 0.01)$ were used to train the MRFResnet encoder $\mathcal{R}$. We created fifty noisy realisation of each fingerprint (i.e. data augmentation), and for each we performed dictionary search to find correct T1/T2 (closest match) training labels and not those that originally generated the fingerprints. Noisy data augmentation creates a tube around the finely-sampled Bloch response manifold, and the label-search procedure enables learning a Euclidean *projection* mapping (rather than a possible overfitted denoiser) onto this manifold. Trainings used Adam optimiser with MSE loss for 20 epochs, 0.01 initial learning rate with decay factors 0.8/0.95 and mini-batch sizes 500/20 for $\mathcal{R}$ and $\mathcal{G}$, respectively, and per epoch 10% of the training samples were held out for validation. The same datasets were used for training KM's encoder and decoder models using LBFGS optimiser.

6.3. Deep vs. shallow learning models' prediction results

To test the prediction performances of the learned models MRFResnet and KM, 500K out-of-sample noisy fingerprints were randomly generated and fed to the corresponding encoder models to estimate the T1/T2 parameters. Predicted

[4]Used hyperparameters: KM used optimised kernel scales by the MATLAB's fitrkernel function and 1000/500 random features (Wilson et al., 2016) per output index for the encoder/decoder models, respectively. VS used 880 shared views as in (Gómez et al., 2020). LRTV used the TV norm weighting $\forall i, \lambda_i = \lambda = 0.2/0.04$ for the 2D/3D brain scans, respectively. AIRMRF used 2D/3D gaussian filters with spreads $\sigma = 1/0.5$ for the 2D/3D brain scans. FLOR used the nuclear norm weighting $\lambda = 10$ for 2D brain scans. Parameters were adjusted experimentally according to the visual impression of the results and where ground-truth was available (e.g. for the phantom or retrospective experiments) they were grid-searched to minimise the reconstruction errors.

	2D spiral		2D radial		3D spiral	
MAPE (%)	T1	T2	T1	T2	T1	T2
ZF-DM	7.36	12.80	9.93	15.46	6.69	16.29
*-MRFResnet	7.49	13.42	9.89	15.79	6.54	16.88
VS-DM	31.46	19.53	28.54	16.04	8.47	**14.43**
*-MRFResnet	33.65	19.85	30.67	16.48	8.53	**14.90**
LR-DM	10.65	15.99	10.75	16.67	6.18	15.04
*-MRFResnet	11.15	16.27	10.94	17.05	5.99	15.25
LRTV-DM	**5.60**	**11.44**	**5.54**	**9.76**	**5.87**	15.29
*-MRFResnet	**6.03**	**11.96**	**5.85**	**10.69**	**5.68**	15.44

Table 2: Average errors for predicting the T1/T2 values of the *in-vitro* phantom ROIs using DM-free reconstruction algorithms ZF, VS, LR and LRTV (ours), followed by DM and MRFResnet (ours) for quantitative inference.

T1/T2s were then fed to the decoder models for generating the corresponding noise-free Bloch responses. The ground-truth (GT) T1/T2s from DM were used to measure encoders' performances based on Mean Absolute (Percentage) Errors MAE $= \mathbb{E}[|\widehat{T}1 - T1^{\text{GT}}|]$ and MAPE $= \mathbb{E}[\frac{|\widehat{T}1-T1^{\text{GT}}|}{T1^{\text{GT}}}]$ (similarly for T2). Corresponding clean fingerprints were used as GT to measure generative model (decoder) predictions based on Normalised-RMSE $= \mathbb{E}\frac{\|\mathcal{G}(\widehat{\Theta})-\mathcal{B}(\Theta^{\text{GT}})\|}{\|\mathcal{B}(\Theta^{\text{GT}})\|}$. Table 1 summarises our results.

6.3.1. Discussion

MRFResnet outperforms KM and achieves reliable predictions for T1/T2 values and Bloch response generation, with about or less than 1% average difference with the DM baseline. KM reports poor T2 and Bloch response estimations for the number of random features used. By increasing this number KM's expressive capacity can improve (e.g. we observed by doubling random features KM's T2 error reduces to about 7%), however by comparing both model sizes at their current configurations, we can deduce the advantage of *depth* in MRFResnet to embed DM more efficiently compared to its shallow alternative KM for the adopted acquisition sequence.

6.4. In-vitro phantom experiment

The 2D (spiral/radial) and 3D (spiral) acquisition schemes were tested for measuring quantitative parameters in twelve tubes of the EUROSPIN TO5 phantom. Table 2 shows the MAPE errors averaged over all region of interests (ROIs/tubes) for predicting the T1/T2 values using different DM-free reconstructions algorithms ZF, LR, VS and the proposed LRTV, all cascaded to DM or the proposed MRFResent for quantitative inference. The gold standard T1/T2 values reported in the phantom's manual were used as reference. DM and MRFResent score very similar quantitative inference accuracies regardless of the reconstruction scheme, which shows that the MRFResnet accurately embeds DM. In addition, Figure 4 displays the mean and standard deviation of the predicted T1/T2 values in each ROI. For compactness we only display ZF-DM, LR-DM and VS-DM baselines, and the proposed LRTV-MRFResnet predictions. Computed parameter map images are also shown in the supplementary materials (Figure S5). Figure 5 displays the Bland-Altman plots of the percentage differences between T1/T2 values of the phantom ROIs in spiral and radial scans, estimated using the ZF-DM and LRTV-MRFResnet.

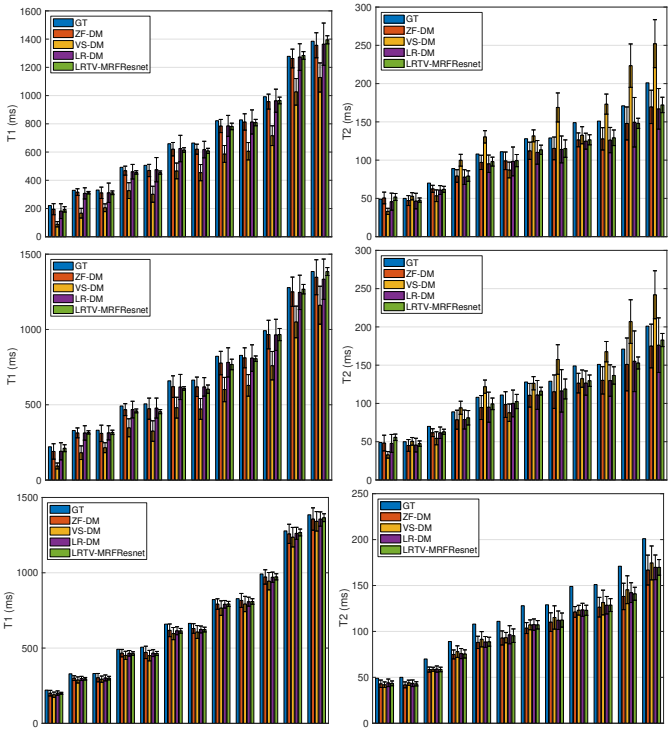


Fig. 4: The mean T1 (left column) and T2 (right column) values in milliseconds and their standard deviations (error bars) estimated via using four reconstruction methods compared to the reference values (GT) in 12 phantom ROIs. Results are compared for 2D spiral (top row), 2D radial (middle row) and 3D spiral acquisitions (bottom row).

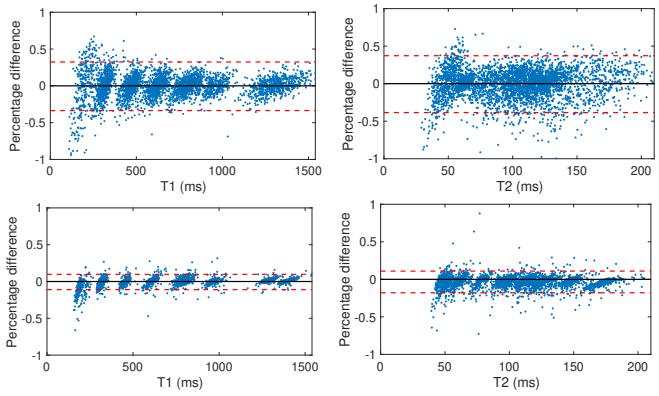


Fig. 5: Confidence intervals (CIs) for the percentage differences between predicted T1/T2 values of the phantom ROIs in 2D spiral and radial scans, using ZF-DM (top) and the LRTV-MRFResnet (bottom) reconstructions. The 0.95%-CIs are -33.9% to 32.3% for T1 and -38.5% to 37.4% for T2 using ZF, whereas using LRTV CIs are tighter -10.2% to 7.8% for T1 and -17.0% to 10.4% for T2.

6.4.1. Discussion

From Figure 4 we observe that tested methods (except VS for 2D) report comparable predictions for the mean T1/T2 values in each ROI.[5] T1 values are comparable to the GT (although ZF, LR and VS slightly underestimate T1). The predicted T2 values, especially in high T2 regimes, are under-estimated (nega-

[5]We observe VS generally trades off image smoothness against overestimated T2s and underestimated T1s. This compromise is strongly unfavourable in 2D acquisitions. Larger k-space neighbourhood information was available/shared in 3D (than 2D) acquisitions, which made 3D VS competitive.

tive bias). We hypothesise this is due to physical effects e.g. flip angle calibration errors, diffusion or magnetisation transfer that are currently un-modelled in the reconstruction schemes. Overall, the proposed LRTV predicts least biased T1/T2s. Notably, LRTV has the *least variations* around the estimated values (see the error bars in Figure 4). For all experiments and averaged over all ROIs, the LRTV's standard deviation is 1.5/2.5 times less than its closest competitor for predicting T1/T2, respectively. This leads the LRTV to have the least (or close to least) predication errors in all tested acquisitions (see Table 2). From Figures 5 and images shown in Figure S5 we observe that LRTV *enables highly consistent predictions* across radial and spiral sampling protocols i.e. per-pixel estimated T1/T2 values in all ROIs obtained from 2D radial and spiral measurements are 2 to 3 times more consistent with each other than those computed via ZF. We did not observed similar consistency level in other tested algorithms as the readout-dependent undersampling artefacts in images were not fully removed by them.

6.5. *In-vivo experiments*

Three sets of experiments were conducted: First, we used the 2D and 3D acquisitions for scanning a healthy volunteer's brain (real-world scans at 1.5T). Figures 6 and S6 display the parametric maps reconstructed from the 2D spiral and radial readouts. We computed the T1, T2 and proton density (PD) maps using baseline reconstruction algorithms ZF, VS, LR, FLOR, AIR-MRF and our proposed LRTV. While baselines use DM either for quantitative inference or also during reconstruction (i.e. AIR-MRF), we further compare the DM-free LRTV's performance when cascaded to DM, KM and MRFResnet for quantitative inference. For the 3D spiral acquisitions we compared LRTV and its closest competitor VS in Figure 7. Outcomes from other tested algorithm are displayed in the supplementary materials (Figure S7). Since FLOR does not use dimensionality-reduction, our system ran out of memory during 3D reconstruction; hence results are not reported in this case. Second, we used a 3D spiral acquisition at a 3T scanner for imaging a glioma patient, where the LRTV-MRFResnet maps was compared to its closest competitor VS-DM in Figure 9, in addition to ZF-DM in Figure S8. Figure S9 also plots the small differences made in paramtric maps by replacing MRFResnet with DM, after LRTV reconstruction. Further, we performed ROI comparisons on *in-vivo* data: For the healthy subject, we obtained brain tissue segmentation maps (RIOs) using the (Zhang et al., 2001) algorithm, and for the glioma patient, ROIs were manually segmented by a trained neuroradiologist (Figure 8). The ROI-based predicted T1/T2 values of the proposed LRTV-MRFResnet were then compared to the LRTV-DM, as well as the baseline VS-DM (also ZF-DM) which for the used acquisition protocols, it was previously validated and shown consistent with the literature values (Gómez et al., 2020; Pirkl et al., 2020; Fujita et al.). Table 3 shows that the differences made by the LRTV-MRFResnet versus MRFResnet-DM and a previously validated baseline VS-DM are small (less than %5/%8 for the T1/T2 values, respectively) across all ROIs and field strengths, indicating methods' reliable quantification for the healthy tissues (white/grey matter) as well as unhealthy

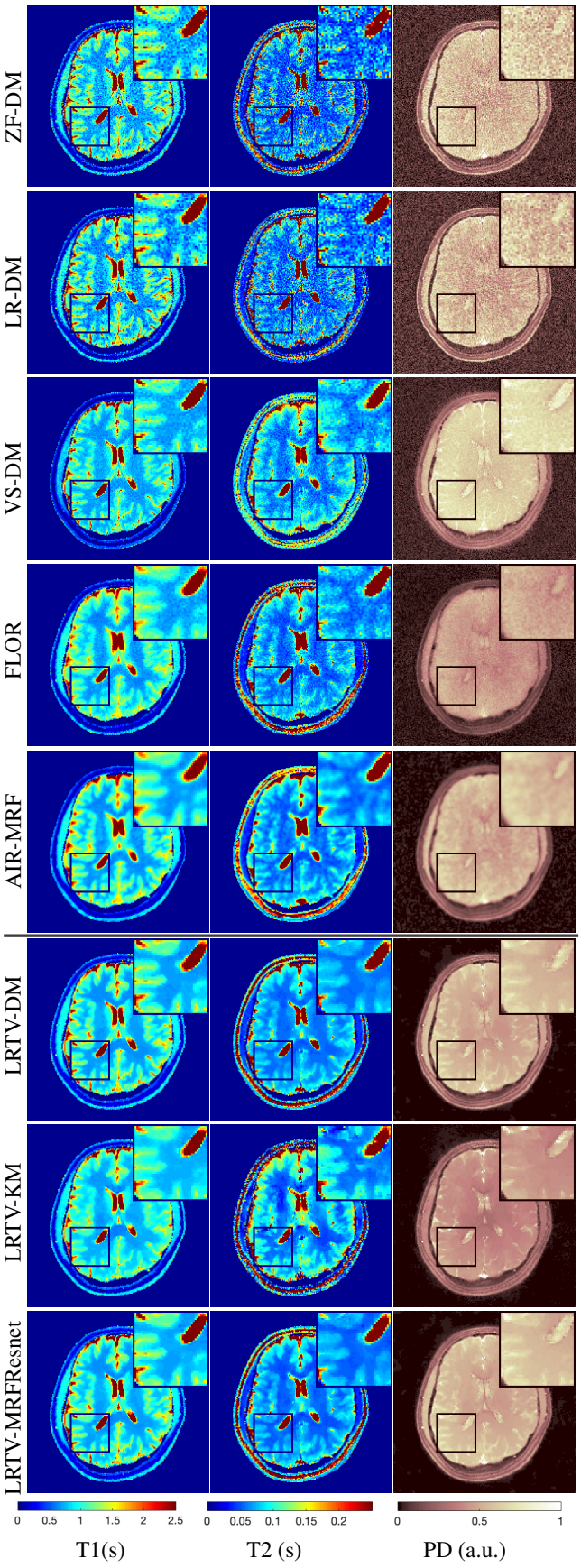


Fig. 6: Reconstructed T1, T2 and PD maps of a healthy volunteer brain from a 2D spiral acquisition using different reconstruction and inference algorithms (real-world scan at 1.5T).

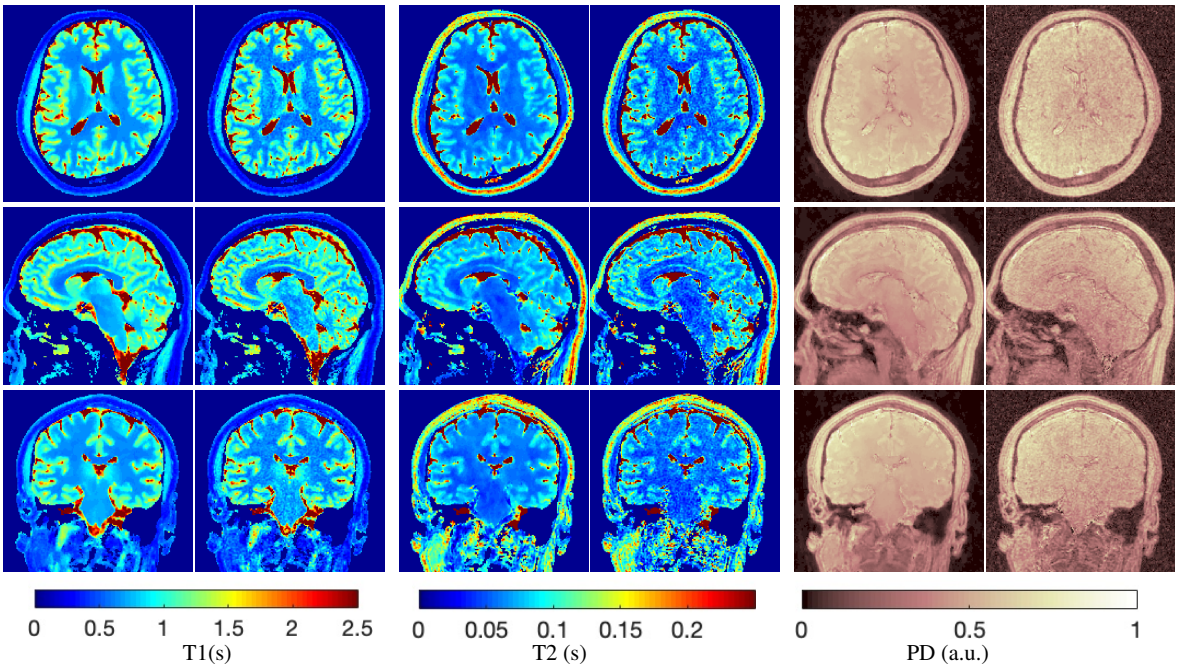


Fig. 7: Reconstructed T1, T2 and PD maps of a healthy volunteer brain using 3D acquisition with spiral readouts (real-world scan at 1.5T). The (zoomed) 3D maps are computed using LRTV-MRFResnet (left sub-column) and the VS-DM (right sub-column) algorithms.

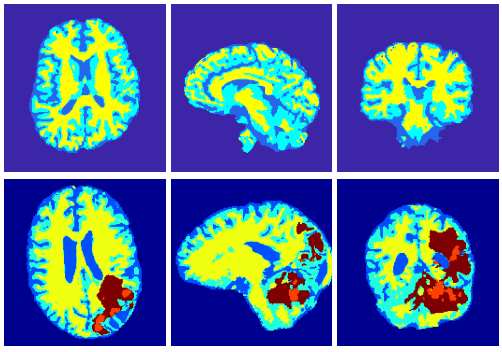


Fig. 8: 3D segmented tissues for the healthy (top) and patient (bottom) subjects: cerebrospinal fluid (light blue), white matter (yellow), grey matter (cyan), peritumoral edema (dark red), solid tumour (light red), background (dark blue).

	White matter		Grey matter		Edema		Tumour	
diff. (%)	T1	T2	T1	T2	T1	T2	T1	T2
Healthy subject								
ZF-DM	5.47	10.15	1.80	1.24	—	—	—	—
VS-DM	4.66	7.85	0.59	2.00	—	—	—	—
LRTV-DM	0.33	0.59	1.83	1.15	—	—	—	—
Glioma patient								
ZF-DM	1.72	8.24	0.99	4.72	0.26	3.32	0.29	3.96
VS-DM	0.82	5.94	0.94	2.54	0.12	2.52	0.29	2.48
LRTV-DM	0.51	5.83	0.39	6.69	0.38	1.07	0.96	2.72

Table 3: The ROI-based difference (%) between the average T1 and T2 values obtained from LRTV-MRFResnet and the methods ZF-DM, VS-DM and LRTV-DM for the healthy (1.5T scan) and glioma patient (3T scan) subjects.

tissues (peritumoral edema, solid tumour) where substantial changes in the brain occur.

Due to the lack of ground truth in the two first experiments, we conducted a third experiment with retrospective validations against the ground-truth anatomical maps. Here the gold standard anatomical maps were acquired from a separate volunteer using MAGIC quantitative imaging protocol (Marcel, 2015). Figure S2 shows these ground truth parameter maps. From these parametric maps we then constructed the corresponding TSMIs and noisy MRF measurements. We used the same excitation sequence and the 2D spiral and radial k-space sampling patterns as in our real-world scans to construct a single-coil acquisition where the k-space measurements were corrupted by additive Gaussian noise with 35 dB SNR. Table 4 compares the reconstruction accuracies of the baselines against our proposed algorithm for the T1, T2, and PD maps as well as the TSMI images. Accuracies were measured by the MAE and MAPE errors, reconstruction SNRs, and the Structural Similarity Index Metric (SSIM) (Wang et al., 2004). Reconstructed maps and their differences with respect to the ground-truth can be found in supplementary Figures S3 and S4.

6.5.1. Discussion

The LRTV-DM and LRTV-MRFResnet perform on par, and both outperform all tested baselines for reconstructing T1, T2 and PD maps in all acquisition schemes. This can be observed both visually in Figures 6, S6, 7,9, S3 and S3, and quantitatively in Table 4 across all tested metrics. Other baselines were unable to successfully remove the under-sampling artefacts in TSMIs, and these errors propagated to the parameter inference phase and resulted in inaccurate maps. Temporal-only priors incorporated within LR are shown insufficient to regularise the inverse problem and LR sometimes (e.g. 2D spiral acquisitions) can admit solutions with even stronger artefacts than the model-free ZF baseline. This issue was previously studied for other non-Cartesian MRF readouts that similar to our spiral/radial trajectories, miss to sample the corners of the k-space in all time-frames (see section 2.2.2 and figure 2 in (Cline et al., 2017)). In the absence of reference for the k-space corners information, the LR iterations despite minimising the objective $\|Y-\mathcal{A}(VX)\|^2$ can converge to solutions with high-frequency artefacts, as visible in the computed maps. This highlights the need for adding

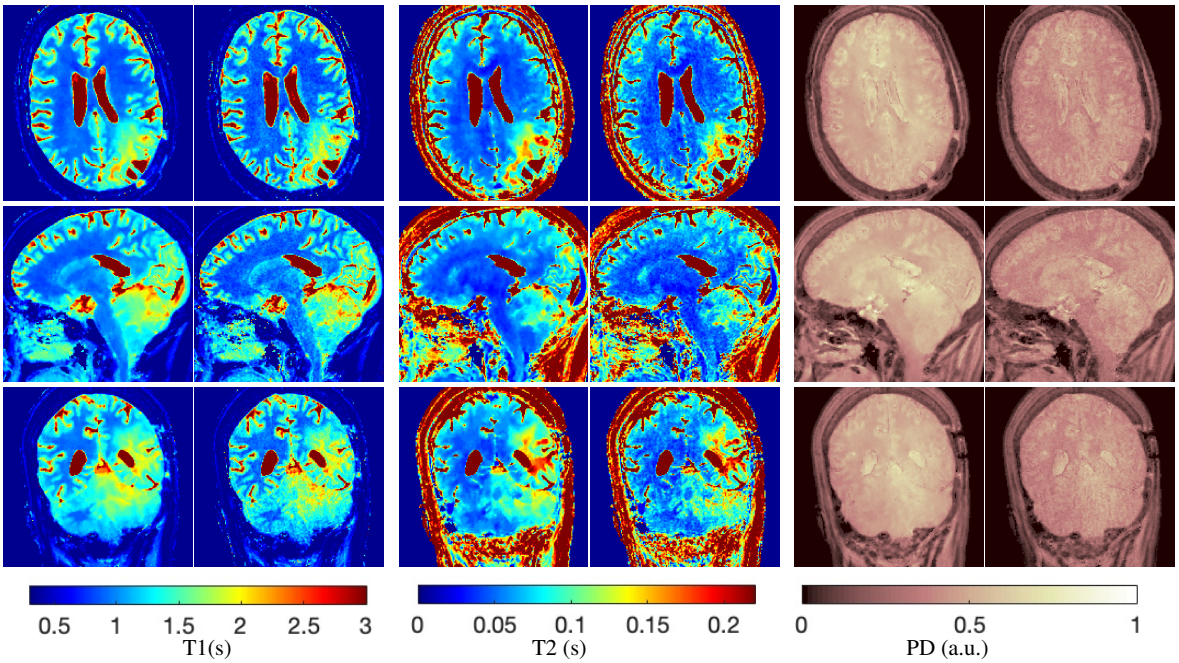


Fig. 9: Reconstructed T1, T2 and PD maps of a glioma patient's brain using 3D acquisition with spiral readouts (real-world scan at 3T). The (zoomed) 3D maps are computed using LRTV-MRFResnet (left sub-column) and the VS-DM (right sub-column) algorithms.

	2D spiral acquisition									2D radial acquisition								
	MAPE (%)		MAE (ms)		SNR (dB)		SSIM			MAPE (%)		MAE (ms)		SNR (dB)		SSIM		
	T1	T2	T1	T2	PD	TSMI	T1	T2	PD	T1	T2	T1	T2	PD	TSMI	T1	T2	PD
ZF-DM	6.58	14.87	64.2	14.7	14.45	14.62	95.59	95.56	88.34	9.90	27.52	97.6	28.1	14.84	11.65	90.34	87.43	89.01
VS-DM	9.19	23.74	99.2	23.8	17.02	19.16	94.40	93.58	90.08	9.32	23.63	99.1	23.8	16.52	18.28	93.40	93.09	89.87
LR-DM	9.96	19.39	98.6	19.5	17.72	13.26	90.15	93.13	90.76	7.49	13.46	82.2	14.1	19.97	15.83	92.96	96.23	95.90
AIRMRF	10.64	15.38	126.1	17.5	15.08	10.98	90.43	94.59	94.99	10.87	16.28	128.7	18.5	14.93	10.72	90.12	93.96	94.72
FLOR	7.69	17.97	86.0	16.8	18.32	21.61	94.98	96.80	97.75	9.85	24.80	115.3	22.7	16.26	19.91	93.72	95.31	96.48
LRTV-DM	**4.06**	**7.52**	**42.3**	**8.3**	**26.59**	**25.36**	**97.21**	**98.61**	**98.93**	**4.90**	**8.73**	**52.8**	**10.0**	**25.15**	**23.79**	**96.32**	**97.91**	**98.64**
*-KM	4.52	16.81	46.5	15.6	15.55	–"–	96.34	95.40	96.53	5.15	18.32	55.0	17.5	15.21	–"–	97.79	94.75	96.33
***-MRFResnet**	**4.11**	**7.94**	**42.9**	**8.6**	**26.23**	–"–	**97.27**	**98.53**	**98.87**	**4.96**	**9.13**	**53.5**	**10.2**	**24.82**	–"–	**96.42**	**97.83**	**98.59**

Table 4: The T1, T2, PD and TSMI reconstruction accuracies (measured by the MAE and MAPE errors, reconstruction SNR, and SSIM metrics) of the baselines and our proposed LRTV-MRFResnet algorithm, validated retrospectively against ground-truth anatomical maps (simulated MRF scan).

an appropriate spatial-domain regularisation. FLOR reduces the LR's artefacts but this improvement is limited because the suggested nuclear norm penalty does not incorporate an explicit spatial regularisation. Further for reducing artefacts, FLOR can introduce an undesirable bias in the computed T1/T2 maps e.g. see error maps in Figures S3 and S3. The non model-based VS baseline incorporates spatial regularisation and results in spatially smoother maps than ZF and LR, but it is unable to output artefact-free images. The model-based AIR-MRF adds spatial regularisation through 2D/3D low-pass Gaussian filters however this trades off the sharpness of the computed maps and can increase the errors at the tissue boundaries (we searched Gaussian spreads that keep the blurs and high-frequency artefacts minimal). For our acquisition readouts, Gaussian filters performed better than disk filters of (Cline et al., 2017) for avoiding strong Gibbs artefacts. On the other hand, the spatiotemporally regularised LRTV greatly improves the TSMI reconstructions i.e. 4 dB enhancement compared to the closest competitor baseline (Table 4). This enables computing accurate and aliased-free multi-parametric inference using DM or the DM-free learning-based alternative MRFResnet as visible in Figures 6, S6, 7, 9, S3 and S4. MRResnet and DM score competitive quantitative inference results i.e. T1 and T2 MAPE less than 5% and 9%, respectively (Table 4). KM also outputs comparably accurate T1 maps, however this shallow learning model despite having a model size larger than MRFResnet, is unable to learn accurate T2/PD quantification and results in poor estimated maps, consistent with observations in section 6.3.

6.6. An ablation study on deep inference models

While MRFResnet is not an integral part of the reconstruction i.e. it can be ablated and replaced by DM or other deep learning baselines, here we show that it is indeed an accurate and efficient method of inference, and it performs on-par with the DM. Here we use the LRTV algorithm for TSMI reconstruction, and compare differences (NRMSE) between the T1 and T2 maps obtained from DM and deep learning inference baselines for the *in-vitro* and *in-vivo* real-world scans. Results are reported in Table 5 and for the *in-vivo* 2D spiral scan are illustrated in Figure S1. The MRFResnet is compared against: the fully-connected neural network (FNN) of (Cohen et al., 2018), the time-domain convolutional network (CNN) of (Hoppe et al.,

NRMSE (%)		T1	T2	T1	T2
Models	#params.	2D/3D volunteer scans		2D/3D phantom scans	
MRFResnet (enc.)	1.3k	3.25 / 1.28	7.15 / 2.68	0.08 / 0.13	0.12 / 0.13
FNN	355.2k	4.12 / 2.91	9.17 / 3.60	0.63 / 0.60	1.82 / 1.05
CNN	58.4k	8.29/ 11.94	36.88 / 28.64	0.65 / 0.74	2.73 / 2.34
CNN*	58.4k	6.65 / 5.14	19.46 / 10.38	1.23 / 1.31	2.01 / 2.13
Residual CNN	3.9M	12.16 / 13.33	31.75 / 16.85	0.99 / 1.30	1.87 / 2.18
Residual CNN*	3.9M	7.53 / 2.60	14.01 / 8.51	1.21 / 0.98	6.22 / 1.85
Subspace FNN	8.3k	2.05 / 1.64	7.19 / 3.07	0.08 / 0.09	0.13 / 0.13

Table 5: Differences (NRMSE) between the T1/T2 maps obtained from dictionary matching (DM) and deep learning methods, after LRTV reconstruction (real-world scans). Deep models are the MRFResnet encoder (posposed), FNN (Cohen et al., 2018), CNN (Hoppe et al., 2017), Residual CNN (Song et al., 2019) and Subspace FNN (Golbabaee et al., 2019a). CNN/ResidualCNN are trained on clean dictionary, and (*) indicate same models trained on noisy dictionary. The number of learnable params (i.e. model-sizes) are also reported.

	reconstruction times (s)					inference times (s)	
	ZF	VS	LRTV	AIRMRF	FLOR	DM	MRFResnet
2D	2.0	1.2e1	2.9e1	1.1e2	6.2e4	8.5	< 0.5
3D	1.5e2	9.1e2	2.2e3	4.7e4	—	1.6e3	54

Table 6: Tested runtimes for quantitative brain image computing (8-coil scans).

2017), the time-domain residual convolutional network (Residual CNN) of (Song et al., 2019), and the Subspace FNN network of (Golbabaee et al., 2019a). Baselines, unlike MRFResnet, estimate only the T1/T2 values and not the PD (the PD consistency of MRFResnet with DM is separately reported in supplementary Table S1). While CNN/ResidualCNN were originally proposed to train on clean MRF dictionaries, for enhancing their robustness on real-world data, we also compared the same models (*) trained on noisy dictionary. Other baselines are trained with noisy dictionary by following the model structures and training instructions provided in the corresponding articles. Except MRFResnet and Subspace FNN that enjoy the subspace dimensionality-reduction, other baselines operate in full temporal dimension which result in much larger and also less accurate models (Table 5, Figure S1). Consistent with section 6.3, we also observe very small differences between DM and MRFResnet's parametric inference. Owning to the subspace dimensionality reduction and its residual structure that enable going layer-wise deep and hierarchically segment the Bloch manifold, MRFResnet achieves similar or better approximation for T1/T2 mapping with a much more compact model (i.e. small number of learnable parameters) compared to the tested deep learning baselines.

6.7. Runtimes

Computations were conducted on a 16-core Intel Xeon Gold CPU, 32 GB RAM and a NVIDIA 2080Ti GPU. Where parallel computing was feasible, we adopted GPU implementation for speedup i.e. in forward/adjoint NUFFT operations[6], the TV shrinkage operator[7], VS, MRFResnet and DM. Table 6 includes computation times of the tested methods for the 2D/3D *in-vivo* experiments. LRTV benefits from momentum-acceleration and takes 7-11 iterations to converge, and its runtime faster than tested iterative schemes i.e. about an order of magnitude faster than the DM-based iterative AIR-MRF with fast kd-tree searches. FLOR has the slowest runtime due to not using the subspace dimensionality-reduction, which makes it also memory-wise non-scalable for our 3D reconstruction experiment. We also observed that the LR method without spatial regularisation makes very slow progress towards its (inaccurate) solution and does not converge within our limit of 30 iterations. This indicates that exploiting additional (spatial) solution structure, despite introducing TV shrinkage computations, has an overall runtime advantage (see e.g.(Chandrasekaran and Jordan, 2013; Tang et al., 2017)) by avoiding extra costly forward/adjoint iterations. The LRTV runs 2-3 times slower than its non-iterative competitor VS for achieving better predictions. DM-based inference methods are order(s) of magnitudes slower than MRFResnet, and therefore the great prediction consistency in both approaches suggests adopting neural inference in favour of runtime.

7. Conclusions

We proposed a two-stage DM-free approach for multi-parametric QMRI image computing based on compressed sensing reconstruction and deep learning. The reconstruction is convex and incorporates efficient spatiotemporal regularisations within an accelerated iterative shrinkage algorithm to minimise undersampling artefacts in the computed TSMI. We proposed MRFResnet, a compact encoder-decoder network with deep residual blocks, in order to embed Bloch manifold projections through multi-scale piecewise affine approximations, and to replace the non-scalable DM baseline for quantitative inference. We demonstrated the effectiveness of the proposed scheme through validations on a novel 2D/3D multi-parametric quantitative acquisition sequence. Further in-vivo validations (e.g. cohort data) are required to confirm repeatably of the results and their usage for clinical applications. Future extensions could address limitations such as motion-artefacts and multi-compartment voxel quantification (Cruz et al., 2019; Duarte et al., 2018) that are currently un-modelled in our pipeline. Further accelerations could be studied through stochastic gradients (Tang et al., 2019) and/or learned proximity operations (Tang et al., 2018) where the proposed scheme could complementarily be adopted for creating accurate labelled parametric maps for training.

8. Acknowledgements

Authors thank Prof. Juan Hernandez-Tamames and Prof. Marion Smits (Radiology and Nuclear Medicine dept, Erasmus MC, University Medical Centre Rotterdam) for providing the patient MRF data which was acquired as part of the work statement B-GEHC-05

References

Arberet, S., Chen, X., Mailhe, B., Nadar, M., Speier, P., 2019. Low rank and spatial regularization model for magnetic resonance fingerprinting. US Patent 2019/0041480.

[6] `https://www.opensourceimaging.org/project/gpunufft/`
[7] `https://epfl-lts2.github.io/unlocbox-html/`

Assländer, J., Cloos, M.A., Knoll, F., Sodickson, D.K., Hennig, J., Lattanzi, R., 2018. Low rank alternating direction method of multipliers reconstruction for mr fingerprinting. Magnetic resonance in medicine 79, 83–96.

Balestriero, R., Baraniuk, R., 2018. A spline theory of deep learning, in: Proceedings of the Intl. Conference on Machine Learning, pp. 374–383.

Balsiger, F., Jungo, A., Scheidegger, O., Carlier, P.G., Reyes, M., Marty, B., 2019. Spatially regularized parametric map reconstruction for fast magnetic resonance fingerprinting. arXiv preprint arXiv:1911.03786 .

Balsiger, F., Konar, A.S., Chikop, S., Chandran, V., Scheidegger, O., Geethanath, S., Reyes, M., 2018. Magnetic resonance fingerprinting reconstruction via spatiotemporal convolutional neural networks, in: International Workshop on Machine Learning for Medical Image Reconstruction, Springer. pp. 39–46.

Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences 2, 183–202.

Benjamin, A.J.V., Gómez, P.A., Golbabaee, M., Mahbub, Z.B., Sprenger, T., Menzel, M.I., Davies, M., Marshall, I., 2019. Multi-shot echo planar imaging for accelerated cartesian mr fingerprinting: an alternative to conventional spiral mr fingerprinting. Magnetic resonance imaging 61, 20–32.

Buonincontri, G., Sawiak, S.J., 2016. Mr fingerprinting with simultaneous b1 estimation. Magnetic resonance in medicine 76, 1127–1135.

Bustin, A., Lima da Cruz, G., Jaubert, O., Lopez, K., Botnar, R.M., Prieto, C., 2019. High-dimensionality undersampled patch-based reconstruction (hd-prost) for accelerated multi-contrast mri. Magnetic resonance in medicine 81, 3705–3719.

Cauley, S.F., Setsompop, K., Ma, D., Jiang, Y., Ye, H., Adalsteinsson, E., Griswold, M.A., Wald, L.L., 2015. Fast group matching for mr fingerprinting reconstruction. Magnetic resonance in medicine 74, 523–528.

Chambolle, A., Pock, T., 2011. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision 40, 120–145.

Chandrasekaran, V., Jordan, M.I., 2013. Computational and statistical tradeoffs via convex relaxation. Proceedings of the National Academy of Sciences 110, E1181–E1190.

Cline, C.C., Chen, X., Mailhe, B., Wang, Q., Pfeuffer, J., Nittka, M., Griswold, M.A., Speier, P., Nadar, M.S., 2017. Air-mrf: Accelerated iterative reconstruction for magnetic resonance fingerprinting. Magnetic resonance imaging 41, 29–40.

Cohen, O., Zhu, B., Rosen, M.S., 2018. MR fingerprinting deep reconstruction network (DRONE). Magnetic resonance in medicine 80, 885–894.

Cruz, G., Jaubert, O., Schneider, T., Botnar, R.M., Prieto, C., 2019. Rigid motion-corrected magnetic resonance fingerprinting. Magnetic resonance in medicine 81, 947–961.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems 2, 303–314.

Davies, M., Puy, G., Vandergheynst, P., Wiaux, Y., 2014. A compressed sensing framework for magnetic resonance fingerprinting. SIAM Journal on Imaging Sciences 7, 2623–2656. doi:10.1137/130947246.

Delalleau, O., Bengio, Y., 2011. Shallow vs. deep sum-product networks, in: Advances in Neural Information Processing Systems, pp. 666–674.

Duarte, R., Repetti, A., Gómez, P.A., Davies, M., Wiaux, Y., 2018. Greedy approximate projection for magnetic resonance fingerprinting with partial volumes. arXiv preprint arXiv:1807.06912 .

Fang, Z., Chen, Y., Liu, M., Xiang, L., Zhang, Q., Wang, Q., Lin, W., Shen, D., 2019a. Deep learning for fast and spatially-constrained tissue quantification from highly-accelerated data in magnetic resonance fingerprinting. IEEE transactions on medical imaging .

Fang, Z., Chen, Y., Nie, D., Lin, W., Shen, D., 2019b. Rca-u-net: Residual channel attention u-net for fast tissue quantification in magnetic resonance fingerprinting, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 101–109.

Fujita, S., Buonincontri, G., Cencini, M., et al., . Repeatability and reproducibility of human brain morphometry using three-dimensional magnetic resonance fingerprinting. Technical Report. Wiley Online Library.

Golbabaee, M., Chen, D., Gómez, P.A., Menzel, M.I., Davies, M.E., 2019a. Geometry of deep learning for magnetic resonance fingerprinting, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7825–7829.

Golbabaee, M., Chen, Z., Wiaux, Y., Davies, M., 2017. Cover tree compressed sensing for fast MR fingerprint recovery, in: IEEE Intl. Workshop on Machine Learning for Signal Processing.

Golbabaee, M., Chen, Z., Wiaux, Y., Davies, M., 2019b. Coverblip: accelerated and scalable iterative matched-filtering for magnetic resonance fingerprint reconstruction. Inverse Problems 36, 015003.

Golbabaee, M., Pirkl, C., Menze, M., Bounincotri, G., Gomez, P., 2019c. Deep MR fingerprinting with total-variation and low-rank subspace priors, in: Proceedings of Intl. Soc. Mag. Res. Med. (ISMRM).

Golbabaee, M., Vandergheynst, P., 2012a. Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery, in: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, IEEE. pp. 2741–2744.

Golbabaee, M., Vandergheynst, P., 2012b. Joint trace/tv norm minimization: A new efficient approach for spectral compressive imaging, in: Image Processing (ICIP), 2012 19th IEEE International Conference on, IEEE. pp. 933–936.

Gómez, P.A., Cencini, M., Golbabaee, M., Schulte, R.F., Pirkl, C., et al., 2020. Rapid three-dimensional multiparametric mri with quantitative transient-state imaging. Scientific reports 10, 1–17.

Gómez, P.A., Molina-Romero, M., Bounincontri, G., Sperl, J.I., K, J.D., Menzel, M.I., Menze, B.H., 2016. Simultaneous parameter mapping, modality synthesis, and anatomical labeling of the brain with mr fingerprinting. MICCAI .

Gómez, P.A., Ulas, C., Sperl, J.I., Sprenger, T., Molina-Romero, M., Menzel, M.I., Menze, B.H., 2015. Learning a Spatiotemporal Dictionary for Magnetic Resonance Fingerprinting with Compressed Sensing. 1st International Workshop on Patch-based Techniques in Medical Imaging. MICCAI .

Hoppe, E., Körzdörfer, G., Würfl, T., Wetzl, J., Lugauer, F., Pfeuffer, J., Maier, A., 2017. Deep learning for magnetic resonance fingerprinting: A new approach for predicting quantitative parameter values from time series. Studies in health technology and informatics 243, 202.

Hoppe, E., Thamm, F., Körzdörfer, G., Syben, C., Schirrmacher, F., Nittka, M., Pfeuffer, J., Meyer, H., Maier, A., 2019. Rinq fingerprinting: Recurrence-informed quantile networks for magnetic resonance fingerprinting, in: Medical Image Computing and Computer Assisted Intervention (MICCAI).

Jaubert, O., Cruz, G., Bustin, A., Schneider, T., Koken, P., Doneva, M., Rueckert, D., Botnar, R., Prieto, C., 2020. Free-running cardiac magnetic resonance fingerprinting: Joint t1/t2 map and cine imaging. Magnetic Resonance Imaging 68, 173–182.

Jaynes, E., 1955. Matrix treatment of nuclear induction. Physical Review 98, 1099.

Jiang, Y., Hamilton, J., Lo, W., Wright, K., et al., 2017. Simultaneous T1, T2 and diffusion quantification using multiple contrast prepared magnetic resonance fingerprinting, in: Proc. Intl. Soc. Mag. Res. Med.

Jiang Y, N., Ma, D., Seiberlich, N., Gulani, V., Griswold, M., 2015. MR fingerprinting using fast imaging with steady state precession (fisp) with spiral readout. Magnetic resonance in medicine 74, 1621–1631.

Lerski, R., De Certaines, J., 1993. Ii. performance assessment and quality control in mri by eurospin test objects and protocols. Magnetic resonance imaging 11, 817–833.

Lustig, M., Donoho, D., Pauly, J.M., 2007. Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic Resonance in Medicine 58, 1182–1195.

Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J., Durek, J., Griswold, M., 2013. Magnetic resonance fingerprinting. Nature 495, 187–192.

Marcel, W., 2015. New technology allows multiple image contrasts in a single scan. GESIGNAPULSE.COM/MR SPRING, 6–10.

Mazor, G., Weizman, L., Tal, A., Eldar, Y.C., 2018. Low-rank magnetic resonance fingerprinting. Medical physics 45, 4066–4084.

McGivney, D.F., Pierre, E., Ma, D., Jiang, Y., Saybasili, H., Gulani, V., Griswold, M.A., 2014. SVD compression for magnetic resonance fingerprinting in the time domain. IEEE transactions on medical imaging 33, 2311–2322.

Montufar, G.F., Pascanu, R., Cho, K., Bengio, Y., 2014. On the number of linear regions of deep neural networks, in: Advances in neural information processing systems, pp. 2924–2932.

Nataraj, G., Nielsen, J., Scott, C., Fessler, J., 2018. Dictionary-free mri perk: Parameter estimation via regression with kernels. IEEE transactions on medical imaging 37, 2103–2114.

Nesterov, Y.E., 1983. A method for solving the convex programming problem with convergence rate o (1/kˆ 2), in: Dokl. akad. nauk Sssr, pp. 543–547.

Oksuz, I., Cruz, G., Clough, J., Bustin, A., Fuin, N., Botnar, R.M., Prieto, C., King, A.P., Schnabel, J.A., 2019. Magnetic resonance fingerprinting using recurrent neural networks, in: IEEE Intl. Symposium on Biomedical Imaging (ISBI), pp. 1537–1540.

Pirkl, C.M., Nunez-Gonzalz, L., Endt, S., et al., 2020. Accelerated 3d multi-

parametric mri in glioma patients: Initial clinical experience. Intl' Soc. Mag. Res. in Medicine (ISMRM) .

Rieger, B., Zimmer, F., Zapp, J., Weingärtner, S., Schad, L.R., 2017. Magnetic resonance fingerprinting using echo-planar imaging: Joint quantification of T1 and T2* relaxation times. Magnetic resonance in medicine 78, 1724–1733.

Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena 60, 259–268.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Song, P., Eldar, Y.C., Mazor, G., Rodrigues, M.R., 2019. Hydra: Hybrid deep magnetic resonance fingerprinting. Medical physics 46, 4951–4969.

Tang, J., Egiazarian, K., Golbabaee, M., Davies, M., 2019. The practicality of stochastic optimization in imaging inverse problems. arXiv preprint arXiv:1910.10100 .

Tang, J., Golbabaee, M., Bach, F., et al., 2018. Rest-katyusha: exploiting the solution's structure via scheduled restart schemes, in: Advances in Neural Information Processing Systems, pp. 429–440.

Tang, J., Golbabaee, M., Davies, M., 2017. Gradient projection iterative sketch for large scale constrained least-squares. Proceedings of the Intl. Conference on Machine Learning 70, 3377–3386.

Tofts, P., 2005. Quantitative MRI of the brain: measuring changes caused by disease. John Wiley & Sons.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of machine learning research 11, 3371–3408.

Virtue, P., Stella, X.Y., Lustig, M., 2017. Better than real: Complex-valued neural nets for mri fingerprinting, in: Image Processing (ICIP), 2017 IEEE International Conference on, IEEE. pp. 3953–3957.

Walsh, D.O., Gmitro, A.F., Marcellin, M.W., 2000. Adaptive reconstruction of phased array mr imagery. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 43, 682–690.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13, 600–612.

Weigel, M., 2015. Extended phase graphs: Dephasing, RF pulses, and echoes-pure and simple. Journal of Magnetic Resonance Imaging 41, 266–295.

Wilson, A.G., Hu, Z., Salakhutdinov, R., Xing, E.P., 2016. Deep kernel learning, in: Artificial Intelligence and Statistics, pp. 370–378.

Wright, K.L., Jiang, Y., Ma, D., Noll, D.C., Griswold, M.A., Gulani, V., Hernandez-Garcia, L., 2018. Estimation of perfusion properties with mr fingerprinting arterial spin labeling. Magnetic resonance imaging 50, 68–77.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. IEEE transactions on medical imaging 20, 45–57.

Zhao, B., Setsompop, K., Adalsteinsson, E., Gagoski, B., Ye, H., Ma, D., Jiang, Y., Ellen Grant, P., Griswold, M.A., Wald, L.L., 2018. Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling. Magnetic resonance in medicine 79, 933–942.

Supplementary Materials

SI. Fast iterative shrinkage algorithm with adaptive step-size to solve LRTV reconstruction (pseudo code)

Algorithm 1 for solving LRTV (10)

Inputs: k-space data Y, forward/adjoint operators $\mathcal{A}/\mathcal{A}^H$, subspace V, initial step-size μ, regularisation parameters $\{\lambda_i\}_{i=1}^s$.
Initialization: $k=1$, $X^1=\mathbf{0}$, $\mu_k=\mu\ \forall k=1,2,\dots$
while stopping criterion = false **do**
  $G = V^H\mathcal{A}^H\left(\mathcal{A}(VX^k)-Y\right)$ #(subspace gradient)
  $\nabla = X^k - \mu_k G$ #(gradient update)
  $Z^k_{(i,.)} = \mathbf{Prox}_{\lambda_i\mu_k}(\nabla_{(i,.)}) \quad \forall i=1,\dots s$ #(total variation shrinkage with threshold $\lambda\mu_k$)
  if $\|Y-\mathcal{A}(VZ^k)\|^2 > \|Y-\mathcal{A}(VX^k)\|^2 + 2\mathrm{Re}\langle G, Z^k-X^k\rangle + \mu_k^{-1}\|Z^k-X^k\|^2$ **then**
    $\mu_k = \mu_k/2$ #(adaptive step-size shrinkage)
    go back to line 6
  else
    $X^{k+1} = Z^k + \left(\frac{k-1}{k+2}\right)(Z^k - Z^{k-1})$ #(momentum acceleration)
    $k = k+1$
return reconstructed subspace MRF image X^{k+1}

SII. Comparison between DM and deep learning inference methods

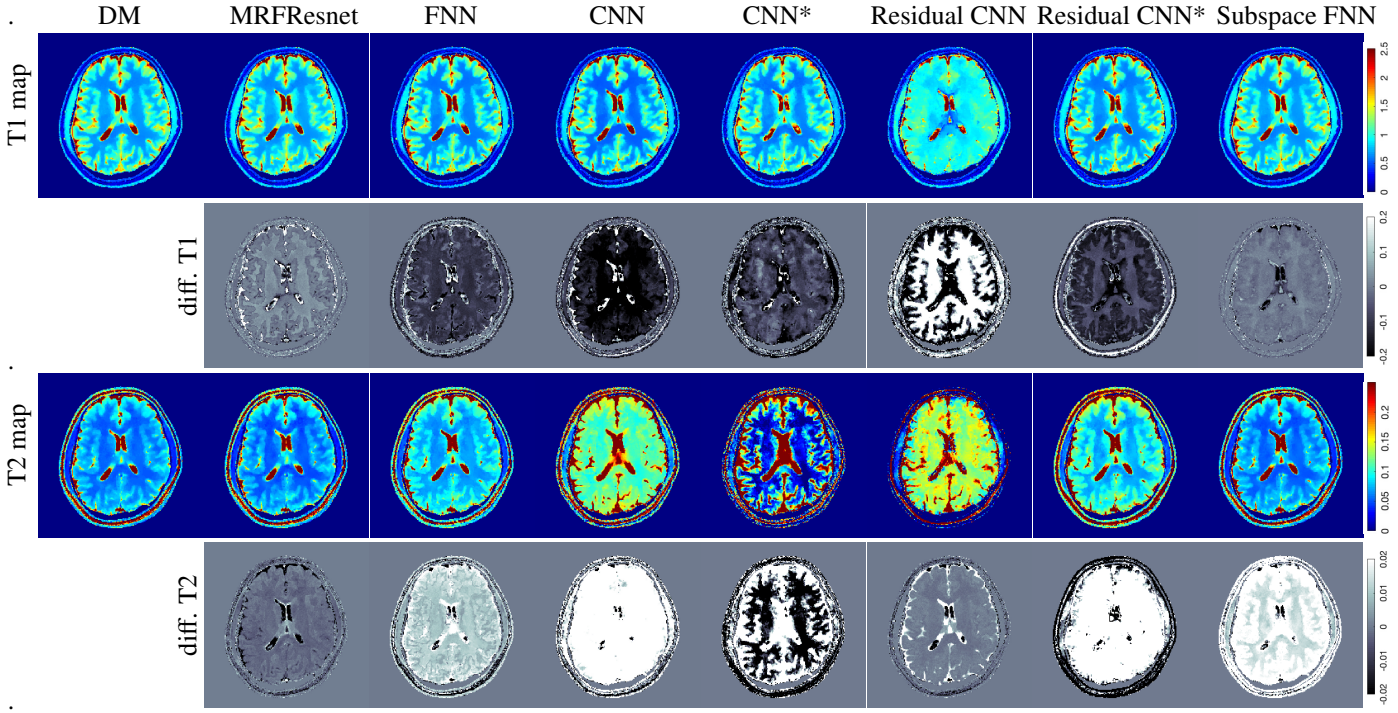


Fig. S1: The computed T1, T2 maps using deep learning baselines and their differences with the maps obtained form dictionary matching (DM), after LRTV reconstruction (healthy subject, real-world *in-vivo* 2D spiral scan at 1.5T). The proposed MRFResnet is compared against the fully-connected neural network (FNN) of (Cohen et al., 2018), the time-domain convolutional network (CNN) of (Hoppe et al., 2017), residual convolutional network (Residual CNN) of (Song et al., 2019), and the (dimension-reduced) Subspace FNN network of (Golbabaee et al., 2019a). CNN/ResidualCNN are trained on clean dictionary, whereas (*) indicate same models trained on noisy dictionary.

SIII. Reconstructed parameter maps for the retrospective experiment

The gold standard anatomical maps were acquired from a volunteer using the MAGIC quantisation protocol (Marcel, 2015), using the 1.5T GE HDxT scanner with 8-channel receive-only head RF coil. Figure S2 shows these ground truth parameter maps. From these parametric maps we constructed the corresponding TSMIs and MRF measurements. A single-coil acquisition with eight times less measurements were considered i.e. $S(\bar{X}) = \bar{X}$ identity sensitivity map. We used the same excitation sequence and the 2D spiral and radial k-space sampling patterns as in our real-world scans (Section 6). The k-space measurements were corrupted by additive i.i.d. Gaussian noise with 35 dB SNR. Figures S3 and S3 show the reconstructed maps and their errors with respect to the ground-truth using various baselines and our proposed LRTV-MRFResnet algorithm.

Results are consistent with those obtained in previous experiments. KM despite great T1 accuracy outputs inaccurate T2/PD predictions i.e. overestimated T2 and underestimated PD maps. Due to the extremely low single coil k-space data for view sharing, VS introduces strong bias on the T1/T2 maps. Further, temporal priors used by LR and FLOR are insufficient to reject under-sampling artefacts, and FLOR can introduce bias in the estimated maps. AIR-MRF's spatial low-pass filtering trades off the sharpness of the computed maps meanwhile unsuccessful to fully remove aliasing artefacts (Figure S3). On the other hand, the spatiotemporally regularised LRTV significantly improves TSMI reconstructions (e.g. 4 dB improvement with respect to the closest competitor baseline, Table 4) through successfully removing strong aliasing artefacts (see Figure S3 and S4). This enables accurate parameter inference in the next stage using DM or the DM-free alternative MRFResnet.

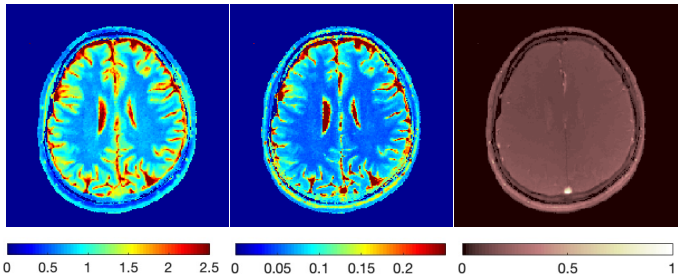


Fig. S2: The T1, T2 and PD maps of a healthy subject brain acquired by the gold standard MAGIC quantitative acquisition protocol.

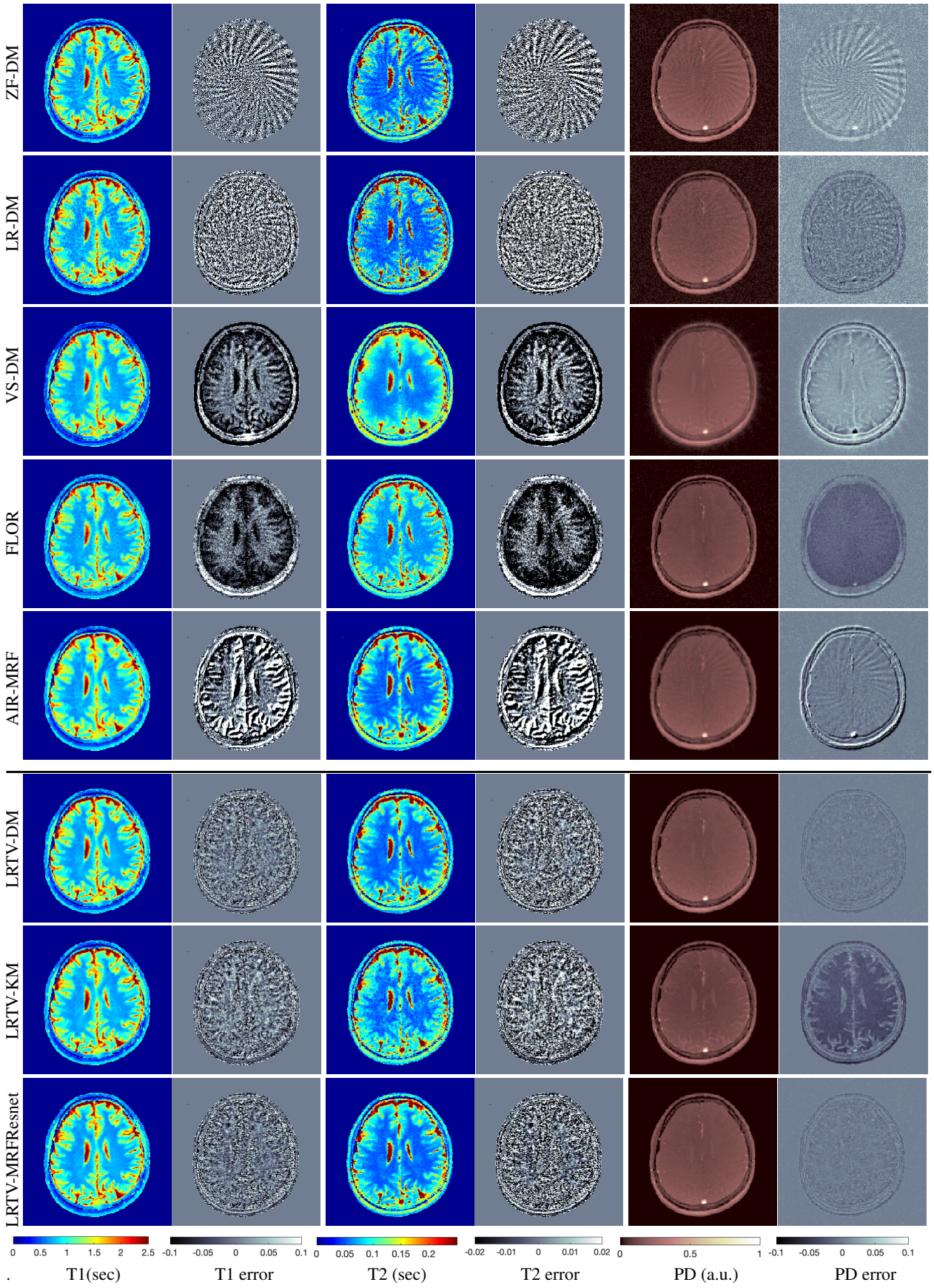


Fig. S3: The computed T1, T2, PD maps of a healthy subject brain and their corresponding errors (with respect to MAGIC gold-standard) using 2D spiral k-space sampling, different reconstruction baselines and our proposed LRTV-MRFResnet algorithm.

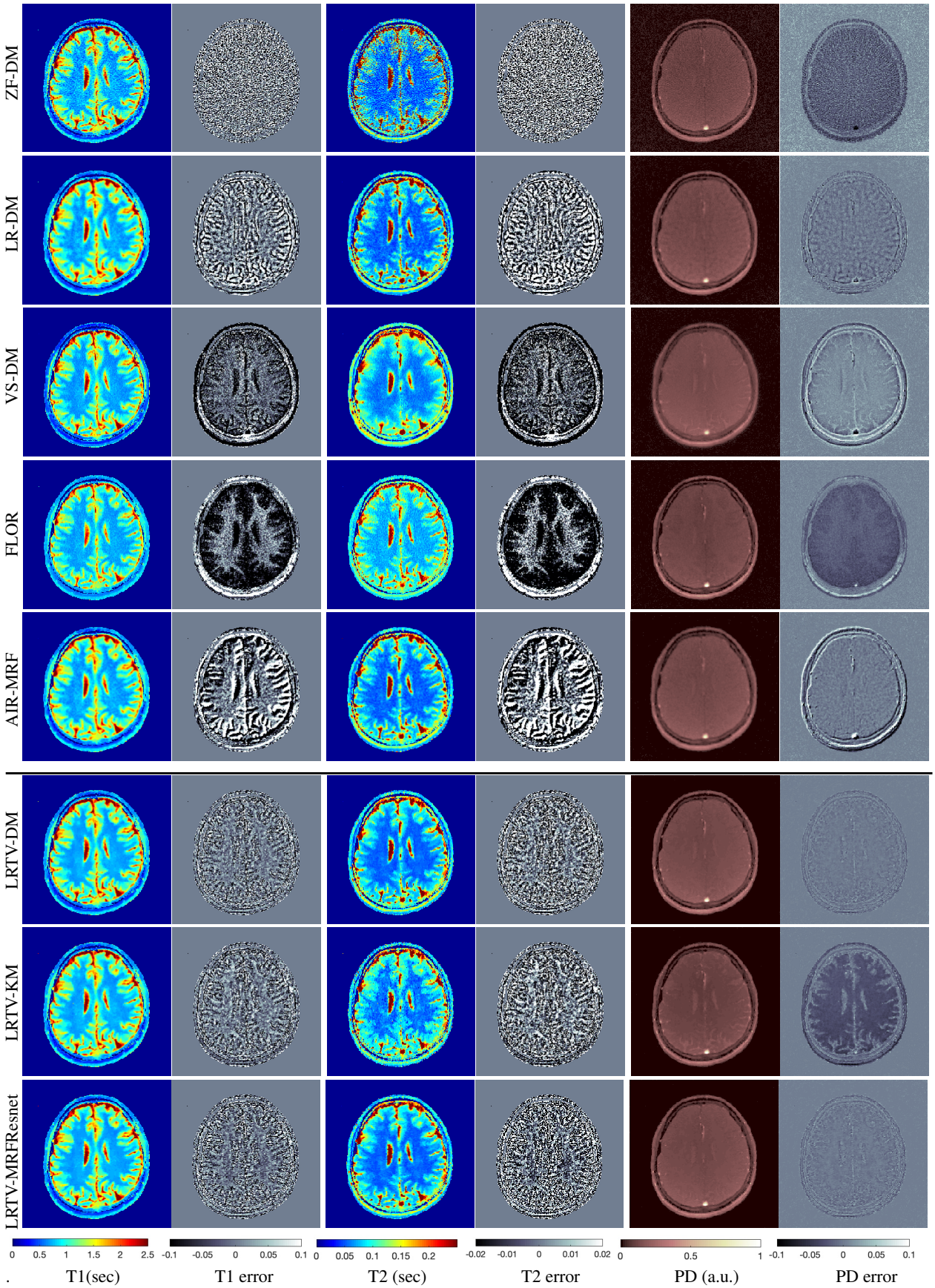


Fig. S4: The computed T1, T2, PD maps of a healthy subject brain and their corresponding errors (with respect to MAGIC gold-standard) using 2D radial k-space sampling, different reconstruction baselines and our proposed LRTV-MRFResnet algorithm.

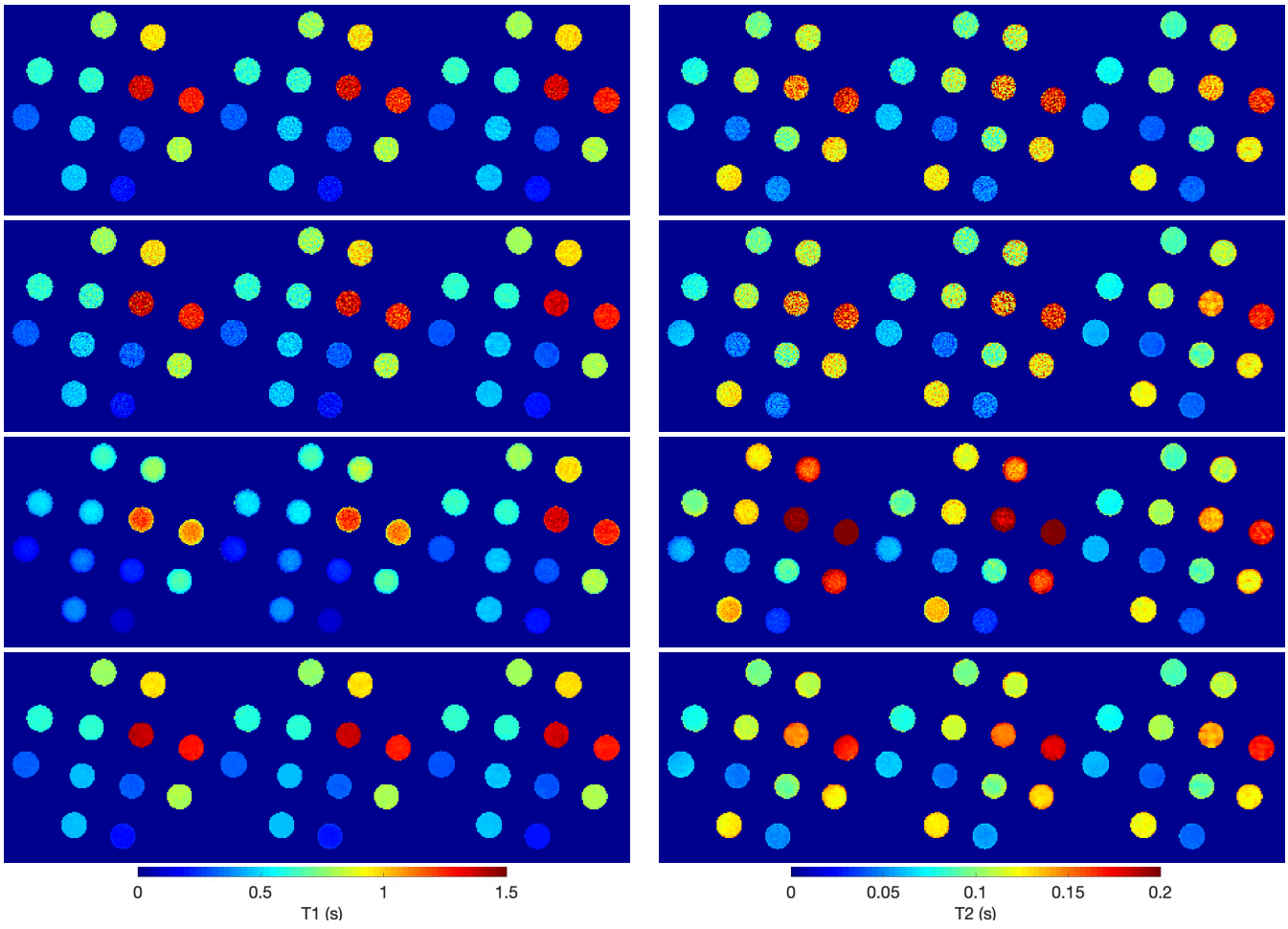


Fig. S5: Reconstructed T1 (3 left columns) and T2 (3 right columns) maps of EUROSPIN TO5 phantom, imaged using the 2D spiral (1st sub-column), 2D radial (2nd sub-column) and 3D spiral (3rd sub-column) k-space acquisitions. Tested methods from top to the bottom row are ZF-DM, LR-DM, VS-DM and the proposed LRTV-MRFResnet algorithm.

SIV. *in-vitro* Phantom reconstructed maps

In Figure S5 we display the computed T1, T2 and PD maps for our *in-vitro* phantom experiments in section VI-D. Tested reconstruction methods are ZF, LR, VS and the proposed LRTV, all fed to the MRFResnet for quantitative inference. Methods ZF and LR result in noisy predictions. It can be observed that for the 2D acquisitions (spiral/radial) VS strongly compromises between outputting smoother images and overestimated T2 values (bias). This issue is also present in *in-vivo* and *in-silico* experiments, where less k-space neighbourhood information are available to share (compared to the 3D acquisitions) and make the VS noncompetitive, and further the overall quantifications inconsistent across 2D/3D acquisitions. The proposed LRTV overcomes this issue through a model-based compressed sensing reconstruction.

SV. Reconstructed maps for the 2D radial *in-vivo* scan

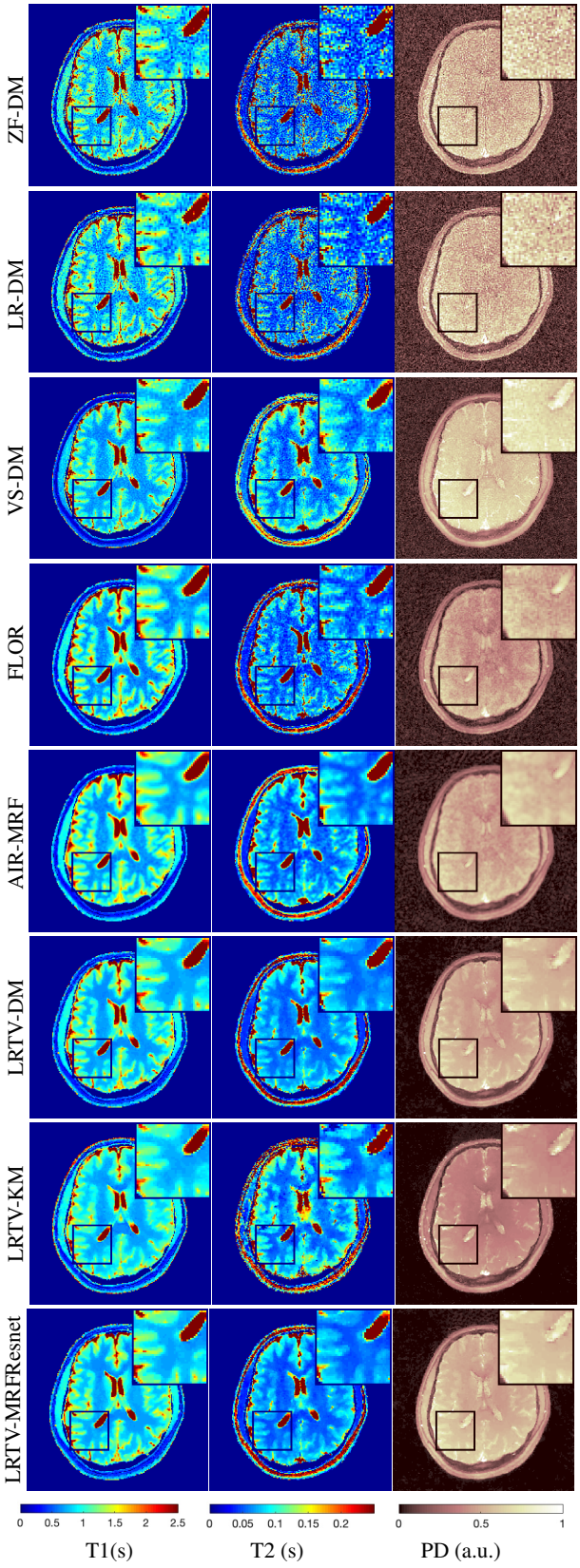


Fig. S6: Reconstructed T1, T2 and PD maps of healthy brain from a 2D radial acquisition using different reconstruction and inference algorithms (real-world scan at 1.5T).

SVI. Supplementary results for the *in-vivo* experiments

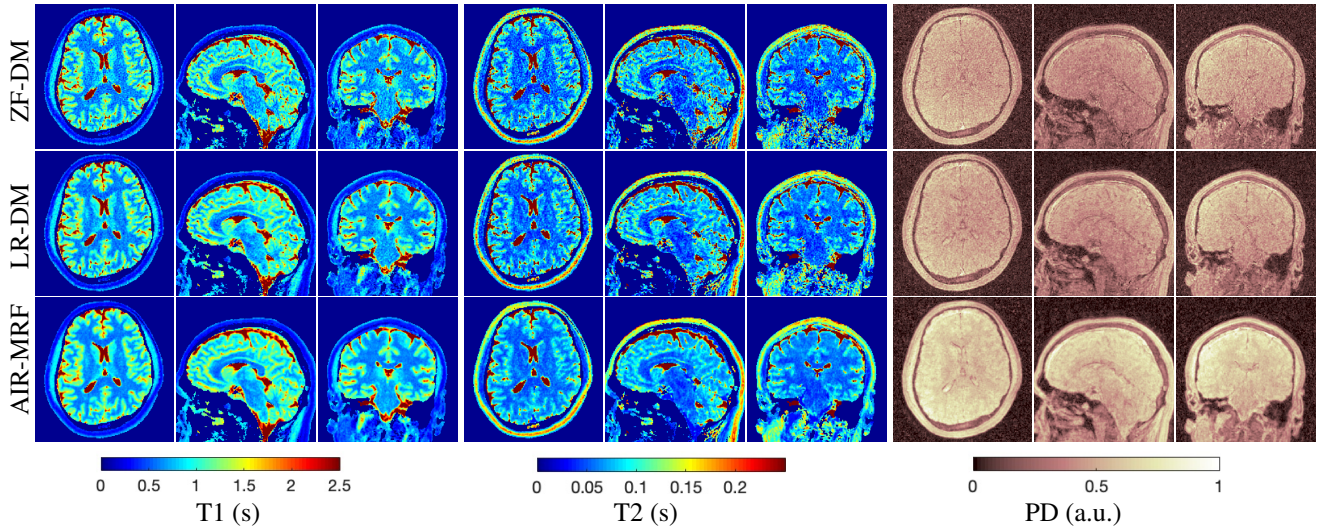


Fig. S7: Reconstructed T1, T2 and PD maps of a healthy brain subject using a 3D scan with spiral readouts (real-world scan at 1.5T). The (zoomed) 3D maps are computed using ZF-DM, LR-DM and AIR-MRF baselines.

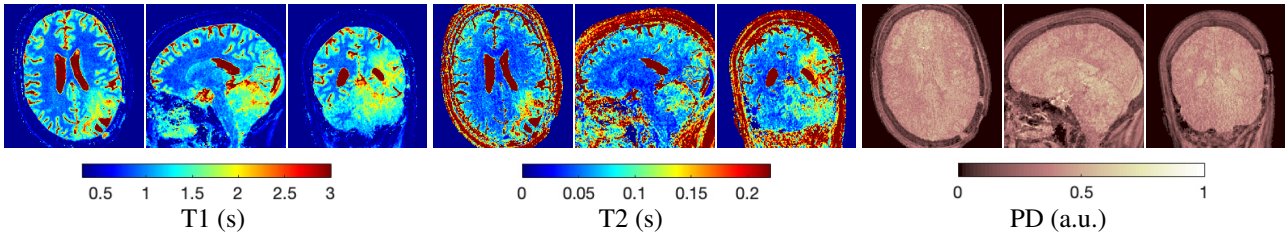


Fig. S8: Reconstructed T1, T2 and PD maps of a glioma patient brain using 3D spiral acquisition and ZF-DM baseline (real-world scan at 3T).

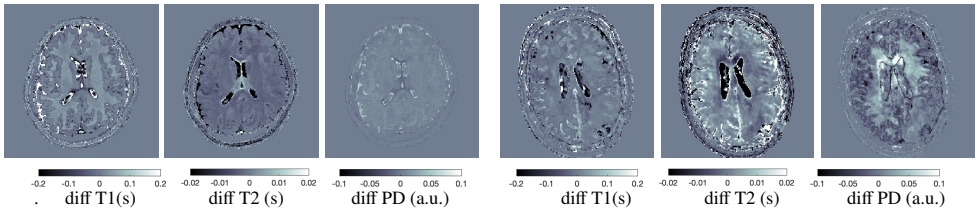


Fig. S9: Differences in the T1, T2 and PD maps between MRFResnet and DM, post LRTV reconstruction: real-world scans of a healthy subject at 1.5T (left) and a glioma patient at 3T (right).

NRMSE (%)	T1	T2	PD
2D/3D phantom scans	0.08 / 0.13	0.12 / 0.13	0.78 / 1.43
2D/3D volunteer scans	3.25 / 1.28	7.15 / 2.68	4.34 / 6.04

Table S1: The NRMSE between the T1, T2 and PD maps obtained from MRFResnet and DM, after LRTV reconstruction (real-world scans).